

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau**BL**

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>5</sup> :</b> <b>C12N 15/12, 15/54, 9/12</b> <b>// C12Q 1/68, C12N 15/11</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 93/15201</b> <b>(43) International Publication Date:</b> <b>5 August 1993 (05.08.93)</b>		
<b>(21) International Application Number:</b> <b>PCT/US93/00586</b> <b>(22) International Filing Date:</b> <b>22 January 1993 (22.01.93)</b>  <b>(30) Priority data:</b> <b>826,935</b> <b>22 January 1992 (22.01.92)</b> <b>US</b>  <b>(60) Parent Application or Grant</b> <b>(63) Related by Continuation</b> <b>US</b> <b>826,935 (CIP)</b> <b>Filed on</b> <b>22 January 1992 (22.01.92)</b>  <b>(71) Applicant (for all designated States except US):</b> <b>NEW ENGLAND DEACONESS HOSPITAL [US/US]; 185 Pilgrim Road, Boston, MA 02215 (US).</b>	<b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only) :</b> <b>AVRAHAM, Hava [IL/US]; 50 Radmor Road, Brighton, MA 02135 (US). GROOPMAN, Jerome [US/US]; 79 Druce Street, Brookline, MA 02146 (US). COWLEY, Sally [GB/GB]; 28 Avonmore Road, London W14 8RS (GB). SCADDEN, David [US/US]; 62 Lexington Street, Weston, MA 02193 (US).</b>  <b>(74) Agents:</b> <b>GRANAHAN, Patricia et al.; Hamilton, Brook, Smith &amp; Reynolds, Two Militia Drive, Lexington, MA 02173 (US).</b>  <b>(81) Designated States:</b> <b>AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</b>  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>			
<b>(54) Title:</b> <b>NOVEL PROTEIN TYROSINE KINASES</b>				
<b>(57) Abstract</b>				
The identification and isolation of novel protein tyrosine kinase genes present on human megakaryocytic and lymphocytic cells, the proteins encoded by these genes, antibodies specific for the encoded proteins, RNA nucleic acid sequences which hybridize to the genes and methods of use therefor.				

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SK	Slovak Republic
CI	Côte d'Ivoire	LJ	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	MC	Monaco	TG	Togo
DE	Germany	MG	Madagascar	UA	Ukraine
DK	Denmark	ML	Mali	US	United States of America
ES	Spain	MN	Mongolia	VN	Viet Nam
FI	Finland				

NOVEL PROTEIN TYROSINE KINASESDescriptionBackground of the Invention

Transduction of signals that regulate cell growth and  
5 differentiation is regulated in part by phosphorylation of  
various cellular proteins. Protein tyrosine kinases are  
enzymes that catalyze this process. Moreover, many act as  
growth factor receptors.

Summary of the Invention

10 The present invention relates to novel protein  
tyrosine kinase genes present in human megakaryocytic and  
lymphocytic cells, the proteins encoded by these genes,  
antibodies specific for the encoded proteins, RNA nucleic  
acid sequences which hybridize to the genes and methods of  
15 use therefor.

The genes isolated as described herein are referred  
to, collectively, as protein tyrosine kinase (pTK) genes.  
The nucleic acid sequences of these genes, isolated as  
discussed herein, show significant homology with  
20 previously identified protein tyrosine kinases containing  
extracellular domains which function as growth factor  
receptors. The pTK genes have been shown to be present in  
both megakaryocytic and lymphocytic cells.

The pTK genes of the present invention show  
25 significant sequence homology with members of the c-kit  
subgroup of growth factor receptors with protein tyrosine  
kinase activity. The c-kit subgroup of receptor tyrosine  
kinases catalyze the phosphorylation of exogenous  
substrates, as well as tyrosine residues within their own  
30 polypeptide chains. (Ullrich, A. and Schlessinger, J.,  
Cell, 61:203 (1990)). Members of the c-kit subgroup  
include FLT/FLK (Fetal Liver Kinase), FGF (Fibroblast

-2-

Growth Factor Receptor) and NGF (Nerve Growth Factor Receptor).

In particular, fourteen pTK genes have been identified. Two pTK genes, referred to as SAL-S1 and SAL-D4 (also referred to as megakaryocyte derived FGF-like receptor tyrosine kinase) were identified in megakaryocytic cells. Five pTK genes, referred to as LpTKs, were identified in lymphocytic cells and have been shown to be present in megakaryocytes as well. One pTK gene, referred to as HpTKs, was identified in human hepatoma cells. Six pTK genes, referred to as bpTK genes, found in human brain tissue.

SAL-S1 is related to the FLT/FLK family of pTKs. SAL-D4 is related to the FGF receptor family of pTKs, and one LpTK (LpTK 3) is related to the NGF receptor family of pTKs.

The pTK genes, which are the subject of the present invention, were identified using two sets of degenerative oligonucleotide primers: a first set which amplifies all pTK DNA segments (SEQ ID NOS:1-2), and a second set which amplifies highly conserved sequences present in the catalytic domain of the c-kit subgroup of pTKs (SEQ ID NOS:3-4). The pTK genes identified in this manner are described below.

SAL-S1 is expressed in several megakaryocytic cell lines, but not in erythroid cell lines. The nucleotide sequence of SAL-S1 was obtained, revealing a sequence containing 158 base pairs. (SEQ ID NO:5). This isolated DNA fragment encoded an amino acid sequence (SEQ ID NO:6) which exhibited significant sequence homology with known protein tyrosine kinases of the FLT/FLK family. The full

-3-

length gene sequence (SEQ ID NO: 17) contains 6827 b.p. and the deduced amino acid sequence (SEQ ID NO: 18) contains 349 residues.

SAL-D4, also expressed in megakaryocytic cells, is a DNA fragment containing the nucleotide sequence of 141 base pairs. (SEQ ID NO:7). This isolated DNA fragment encoded an amino acid sequence (SEQ ID NO:8) which exhibited significant sequence homology with known protein tyrosine kinases of the FGF receptor family.

The LpTKs, including LpTK 2, LpTK 3, LpTK 4, and LpTK 13 and LpTK 25, are expressed in lymphocytic cells, as well as megakaryocytic cells. The nucleotide sequence (151 base pairs) of the LpTK 3 gene was obtained (SEQ ID NO:11), and exhibited significant homology with known protein tyrosine kinases of the NGF receptor family. The nucleotide sequences of the LpTK 2, LpTK 4, and LpTK 13 genes contained 149 base pairs (SEQ ID NO:9), 137 base pairs (SEQ ID NO:13), and 211 base pairs (SEQ ID NO:15) respectively. LpTK 25 has a nucleotide sequence of 3120 b.p. (SEQ ID NO: 22). A full length gene sequence has been obtained for LpTK 2 (SEQ ID NO: 19) which contains 7606 b.p. Additional sequencing of LpTK 4 revealed a sequence of 404 b.p. (SEQ ID NO: 21).

The HpTK 5 gene, expressed in human hepatoma cells, has a nucleotide sequence of 3120 b.p. (SEQ ID NO: 22). Nucleotide sequences of the bpTK's, including bpTK 1, bpTK 2, bpTK 3, bpTK 4, bpTK 5 and bpTK 7 are expressed in human brain tissue encode proteins having the amino acid sequences of SEQ ID NOS: 25-30 respectively.

Thus the present invention includes DNA isolated from a human megakaryocytic cell line, which hybridizes to a

-4-

DNA fragment which hybridizes to DNA encoding an amino acid sequence which is highly conserved in the catalytic domain of protein tyrosine kinases of the c-kit subgroup.

The present invention also includes the proteins  
5 encoded by the pTK genes identified as described herein, which exhibit significant sequence homology with members of the c-kit subgroup of pTKs (i.e. FLT/FLK (SAL-S1), FGF receptor (SAL-D4) or NGF receptor (LpTKS)) as well as the proteins encoded by HpTK 5 and the bpTKs. The present  
10 invention also includes SAL-S1, SAL-D4, and LpTK, HpTK and bpTK homologues or equivalents (i.e., proteins which have amino acid sequences substantially similar, but not identical, to that of SAL-S1, SAL-D4, the LpTKs HpTK and the bpTKs, which exhibit tyrosine kinase activity.) This  
15 invention further includes peptides (SAL-S1, SAL-D4, LpTK, HpTK and bpTK fragments) which retain tyrosine kinase activity, yet are less than the entire SAL-S1, SAL-D4, LpTK, HpTK and bpTK sequences), monoclonal and polyclonal antibodies specific for SAL-S1, SAL-D4, the LpTKs, HpTK  
20 and the bpTKs, and uses for the SAL-S1, SAL-D4, the LpTK, HpTK and the bpTK nucleic acid sequences and SAL-S1, SAL-D4, LpTK, HpTK and bpTK equivalents.

The present invention further includes nucleic acid sequences which hybridize with DNA or RNA encoding the  
25 proteins described herein, which exhibit significant sequence homology with the FLT/FLK, FGF receptor or NGF receptor family of protein tyrosine kinases contained within the c-kit subgroup. Such nucleic acid sequences are useful as probes to identify pTK genes in other  
30 vertebrates, particularly mammals, and in other cell types. They can also be used as anti-sense

-5-

oligonucleotides to inhibit protein tyrosine kinase activity, both in vitro and in vivo.

The SAL-S1, SAL-D4, LpTK, HpTK and bpTK, tyrosine kinases of the present invention can be used as target  
5 proteins in conjunction with the development of drugs and therapeutics to modulate cell growth, differentiation and other metabolic functions. The SAL-S1, SAL-D4, LpTK, HpTK or bpTK proteins can be used as agonists or antagonists to other tyrosine kinases. The SAL-S1, SAL-D4, LpTK, HpTK or  
10 bpTK tyrosine kinases can also be instrumental in the modulation of megakaryocyte and/or platelet adhesion interactions.

In addition, the SAL-S1, SAL-D4, LpTK, HpTK and bpTK tyrosine kinases can be used in screening assays to detect  
15 cellular growth and/or differentiation factors. Using standard laboratory techniques, the ligands of the pTKs of the present invention can be identified. Once identified, assays can be designed to detect these ligands present endogenously, within cells, as well as exogenously, in  
20 extra cellular fluids. Assays can also be designed as diagnostic aids to detect these ligands in body fluids such as blood and urine.

#### Brief Description of the Drawings

Figure 1 depicts the nucleotide sequence of SAL-S1  
25 (SEQ ID NO: 5) and the deduced amino acid sequence (SEQ ID NO:6).

Figure 2 depicts the nucleotide sequence of SAL-D4 (SEQ ID NO:7) and its deduced amino acid sequence (SEQ ID NO:8).

-6-

Figure 3A depicts the nucleotide sequence (SEQ ID NO:9) and its deduced amino acid sequence (SEQ ID NO:10) for LpTK 2.

Figure 3B depicts the nucleotide sequence (SEQ ID NO:11) and its deduced amino acid sequence (SEQ ID NO:12) for LpTK 3.

Figure 3C depicts the nucleotide sequence (SEQ ID NO:13) and its deduced amino acid sequence (SEQ ID NO:14) for LpTK 4.

Figure 3D depicts the nucleotide sequence (SEQ ID NO:15) and its deduced amino acid sequence (SEQ ID NO:16) for the LpTK 13.

Figure 4A-4J depicts the full-length nucleotide sequence (SEQ ID NO: 17) and its deduced amino acid sequence (SEQ ID NO: 18) for SAL-S1.

Figure 5A-5J depicts the full length nucleotide sequence (SEQ ID NO: 19) and the deduced amino acid sequence (SEQ ID NO: 20) for LpTK2.

Figure 6 depicts the partial nucleotide sequence (SEQ ID NO: 21) for LpTK4.

Figure 7A-7D depicts the full length nucleotide sequence (SEQ ID NO: 22) for LpTK25.

Figure 8A-8F depicts the full length nucleotide sequence (SEQ ID NO: 23) and the deduced amino acid sequence (SEQ ID NO: 24) for HpTK5.

Figure 9 depicts the amino acid sequence (SEQ ID NO: 25) of bpTK1.

Figure 10 depicts the amino acid sequence (SEQ ID NO: 26) of bpTK2.

Figure 11 depicts the amino acid sequence (SEQ ID NO: 27) of bpTK3.



-7-

Figure 12 depicts the amino acid sequence (SEQ ID NO: 28) of bpTK4.

Figure 13 depicts the amino acid sequence (SEQ ID NO: 29) of bpTK5.

5 Figure 14 depicts the amino acid sequence (SEQ ID NO: 30) of bpTK7.

#### Detailed Description of the Invention

Novel protein tyrosine kinase genes have been identified, their nucleic acid sequences determined, and  
10 the amino acid sequences of the encoded proteins deduced. The genes isolated as described herein are referred to, collectively, as protein tyrosine kinase (pTK) genes. The nucleic acid sequences of these genes, isolated as discussed herein, show significant homology to  
15 with previously identified protein tyrosine kinases containing extracellular domains which function as growth factor receptors. These genes have been shown to be present in both megakaryocytic and lymphocytic cells.

To facilitate the isolation and identification of  
20 these novel pTKs, two sets of DNA probes were used, as described in the Exemplification. The first set consisted of two degenerative oligonucleotide sequences, pTK 1 (SEQ ID NO:1) and pTK 2 (SEQ ID NO:2) (Matthews, W. Cell 65: 1143 (1991; Wilks, A. F. Proc. Natl. Acad. Sci. USA  
25 86:1603 (1989)). These sequences were used as primers in a polymerase chain reaction to amplify tyrosine kinase DNA segments. (Mullis, K. et al., Cold Spring Harbor Symp. Advan. Biol. 51:263 (1986).

The second set consisted of two oligonucleotide  
30 sequences, pTK 3 (SEQ ID NO:3) and pTKKW (SEQ ID NO:4)

-8-

designed to amplify the nucleic acid sequence which encodes the highly conserved regions of the catalytic domains of the c-kit family of protein tyrosine kinases. These sequences were used as primers in the polymerase chain reaction in a second round of DNA amplification. Using this two-step amplification procedure, DNA fragments which hybridized to these pTK primers were identified, isolated and subsequently sequenced.

In particular, fourteen pTK genes exhibiting significant homology with the c-kit subgroup of protein tyrosine kinases have been identified. Two pTK genes, referred to as SAL-S1 and SAL-D4 (also referred to as megakaryocyte derived FGF-like receptor) were identified in several megakaryocytic cell lines, including CMK 11-5, DAMI, UT-7 and UT-7 grown in erythropoietin, but not in the erythroid cell lines HEL, PMA stimulated HEL cells, or K562. Five pTK genes, referred to as LpTKs, were identified in lymphocytic, as well as in megakaryocytic cells. One pTK gene, referred to as HpTK5 was identified in human hepatoma cells and six genes, referred to as bpTKs, were identified in human brain tissue.

SAL-S1 (SEQ ID NO:6 and 18) encoded by the nucleic acid sequence SEQ ID NOS:5 and 17, exhibits significant homology with the FLT/FLK family of pTKs. SAL-D4 (SEQ ID NO:8) encoded by SEQ ID NO:7, is related to the FGF receptor family of pTKs, and one LpTK (LpTK 3 (SEQ ID NO:12) encoded by the SEQ ID NO:11, is related to the NGF receptor family of pTKs. The remaining LpTKs, LpTK2 (SEQ ID NO:10) encoded by SEQ ID NO:9; LpTK4 (SEQ ID NO:14) encoded by SEQ ID NO:13; LpTK13 (SEQ ID NO:16) encoded by SEQ ID NO:15 LpTK25 encoded by SEQ ID NO: 22, also exhibit

-9-

sequence homology with known protein tyrosine kinases  
(Data not shown).

HpTK5 (SEQ ID NO: 24) encoded by SEQ ID NO: 23 and  
the bpTKs 1, 2, 3, 4, 5 and 7 (SEQ ID NOS: 25-30  
5 respectively, also exhibit sequence homology with known  
protein tyrosine kinases.

Thus, as described above, DNA which hybridize with  
DNA encoding amino acid sequences present in the catalytic  
domain of a protein tyrosine kinase of the c-kit subgroup  
10 of protein kinases have been isolated and sequenced.  
These isolated DNA sequences, collectively referred to as  
pTKs genes, (and their deduced amino acid sequences) have  
been shown to exhibit significant sequence homology with  
known members of receptor tyrosine kinase families.

15 Once isolated, these DNA fragments can be amplified  
using known standard techniques such as PCR. These  
amplified fragments can then be cloned into appropriate  
cloning vectors and their DNA sequences determined.

These DNA sequences can be excised from the cloning  
20 vectors, labeled with a radiolabeled nucleotide such as  
<sup>32</sup>P and used to screen appropriate cDNA libraries to  
obtain the full-length cDNA clone.

The pTk genes as described above have been isolated  
from the source in which they occur naturally, i.e.  
25 megakaryocyte and lymphocytic cells. The present invention  
is intended to include pTk genes produced using genetic  
engineering techniques, such as recombinant technology, as  
well as pTk genes that are synthesized chemically.

The deduced amino acid sequences of the pTk genes  
30 include amino acid sequences which encode peptides  
exhibiting significant homology with the catalytic domain

-10-

of protein tyrosine kinases of the c-kit subgroup of tyrosine kinases. These proteins, encoded by the pTk genes, can include sequences in which functionally equivalent amino acid residues are substituted for  
5 residues within the sequence, resulting in a silent change, that is a change not detected phenotypically. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent,  
10 resulting in a silent substitution.

In addition, the protein structure can be modified by deletions, additions, inversion, insertions or substitutions of one or more amino acid residues in the sequence which do not substantially detract from the  
15 desired functional tyrosine kinases properties of the peptide.

Modified pTKs of the present invention, with receptor tyrosine kinase activity can be made using recombinant DNA techniques, such as excising it from a vector containing a  
20 cDNA encoding such a protein, or by synthesizing DNA encoding the desired protein mechanically and/or chemically using known techniques.

An alternate approach to producing the pTKs of the present invention is to use peptide synthesis to make a  
25 peptide or polypeptide having the amino acid sequence of such a protein. The peptides or modified equivalents thereof, can be synthesized directly by standard solid or liquid phase chemistries for peptide synthesis.

Preferably, the pTKs of the present invention will be  
30 produced by inserting DNA encoding the proteins into an appropriate vector/host system where it will be expressed.

-11-

The DNA sequences can be obtained from sources in which they occur naturally, can be chemically synthesized or can be produced using standard recombinant technology.

This invention also pertains to an expression vector  
5 comprising a pTK gene of the present invention, encoding for a protein which exhibits receptor tyrosine kinase activity.

The pTK genes of the present invention can be used for a number of diagnostic and therapeutic purposes. For  
10 example, the nucleic acid sequences of the pTK genes can be used as probes to identify other protein tyrosine kinases present in other cell types, including eukaryotic and prokaryotic cell types.

The nucleic acid sequences can be used to design  
15 drugs that directly inhibit the kinase activity of protein tyrosine kinases, or to design peptides that bind to the catalytic domain of tyrosine kinases, thus inhibiting their activity. These sequences can also be used to design anti-sense nucleotides that can also inhibit, or  
20 destroy, tyrosine kinase activity. Such inhibition of tyrosine kinase activity would be desirable in pathological states where decreased cellular proliferation would be beneficial, such as leukemias or other malignancies.

25 The nucleic acid sequences can also be used to design drugs, peptides or anti-sense nucleotides as above, but with enhancing, rather than inhibitory effects, on tyrosine kinases. Such enhanced tyrosine kinase activity would result in increasing the phosphorylation of  
30 substrates (exogenous, as well as endogenous tyrosine residues). Enhanced effects would be desirable in states

-12-

where increased cellular proliferation would be beneficial, such as anemias, bleeding disorders and during surgical procedures.

The pTK genes of the present invention can also be  
5 used to obtain soluble fragments of receptor tyrosine kinases, capable of binding their respective ligands (i.e. fibroblast growth factor).

pTK genes encoding soluble receptor tyrosine kinase fragments can be produced using recombinant DNA techniques  
10 or synthetically. In either case, the DNA obtained encodes a soluble pTK fragment which lacks a substantial portion of the hydrophobic transmembrane region to permit solubilization of the fragment.

These soluble pTK protein fragments can be introduced  
15 exogenously to act as competitors with the endogenous, membrane bound pTK for their respective ligands, thus inhibiting tyrosine kinase activity. Alternately, a modified soluble pTK protein fragment can be introduced which binds the ligand but does not activate kinase  
20 activity.

These soluble pTK protein fragments can also be used in binding assays to detect ligands such as growth and differentiation factors. Once these ligands are identified, they may be altered or modified to inhibit or  
25 enhance kinase activity. For example, the ligands may be modified or attached to substances that are toxic to the cell, such a ricin, thus destroying the target cell. The substance may be a super-activating substance which, after binding to the pTK, may substantially increase the kinase  
30 activity, or activate other growth factors.

-13-

PTK genes of the present invention would also be useful to develop diagnostic tools for in vitro screening assays for ligands such as growth factors or differentiation factors that inhibit or enhance kinase activity. The proteins encoded by the pTK genes can also be used in such assays, or as immunogens to produce monoclonal or polyclonal antibodies to be used in such assays.

Such antibodies can also be used in methods of treating conditions in which an individual would benefit therapeutically if protein tyrosine kinase activity could be modified, such as increasing platelet production in bleeding disorders.

The present invention will now be illustrated by the following Exemplification, which is not intended to be limiting in any way.

Exemplification: The Identification and Isolation of the pTK Genes

To facilitate the isolation and identification of these novel pTK genes, two sets of DNA probes were used. (See Table).

The first set consisted of two degenerative oligonucleotide sequences, pTK 1 (SEQ ID NO:1) and pTK 2 (SEQ ID NO:2). These sequences were used as polymerase chain reaction (PCR) primers, using standard PCR techniques, to amplify tyrosine kinase DNA segments.

The second set consisted of two oligonucleotide sequences, pTK 3 (SEQ ID NO:3) and pTKKW (SEQ ID NO:4) selected from the highly conserved regions of the catalytic domains of the c-kit subgroup of protein

-14-

tyrosine kinases. These sequences were also used as polymerase chain reaction primers in a second round of DNA amplification. Using this two-step amplification procedure, DNA fragments which hybridized to these pTK  
5 primers were identified, isolated and subsequently sequenced using known laboratory techniques.

TABLEFirst Round of Amplification

PTK1

10 CGGATCCACAGNGACCT

PTK2

GGAATTCCAAAGGACCAGACGTC

Second Round of Amplification

PTK3 (kit family specific)

15 CGGATCCATCCACAGAGATGT

PTKKW (kit family specific)

GGAATTCCTTCAGGAGCCATCCACTT

Equivalents

Those skilled in the art will recognize, or be able  
20 to ascertain using no more than routine experimentation,  
many equivalents to the specific embodiments of the  
invention described herein. Such equivalents are intended  
to be encompassed by the following claims.



-15-

CLAIMS

The invention claimed is:

1. Isolated DNA of human megakaryocytic origin which hybridizes to a DNA fragment which hybridizes to DNA  
5 encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.
2. Isolated DNA of Claim 1 having a nucleotide sequence selected from the group of nucleotide sequences  
10 consisting of:
  - a) SAL-S1 (SEQ ID NOS:5 and 7);
  - b) SAL-D4 (SEQ ID NO:7);
  - c) LpTK 2 (SEQ ID NOS:9 and 19);
  - d) LpTK 3 (SEQ ID NO:11);
  - 15 e) LpTK 4 (SEQ ID NOS:13 and 21);
  - f) LpTK 13 (SEQ ID NO:15);
  - g) LpTK 25 (SEQ ID NO: 22 and
  - h) HpTK 5 (SEQ ID NO: 23).
3. Isolated DNA of Claim 1 which encodes an amino acid  
20 sequence selected from the group consisting of:
  - a) SAL-S1 (SEQ ID NOS:6 and 18);
  - b) SAL-D4 (SEQ ID NO:8);
  - c) LpTK 2 (SEQ ID NOS:10 and 20);
  - d) LpTK 3 (SEQ ID NO:12);
  - 25 e) LpTK 4 (SEQ ID NO:14);
  - f) LpTK 13 (SEQ ID NO:16);
  - g) HpTK 5 (SEQ ID NO:24);

-16-

- h) bpTK 1 (SEQ ID NO:25);
  - i) bpTK 2 (SEQ ID NO:26);
  - j) bpTK 3 (SEQ ID NO:27);
  - k) bpTK 4 (SEQ ID NO:28);
  - 5 l) bpTK 5 (SEQ ID NO:29); and
  - m) bpTK 7 (SEQ ID NO:30).
4. Isolated DNA of human megakaryocytic origin which comprises a DNA fragment whose sequence encodes an amino acid sequence present in the catalytic domain
- 10 of a protein tyrosine kinase of the c-kit subgroup of protein tyrosine kinases.
5. Isolated DNA of Claim 4 which encodes an amino acid sequence selected from the group consisting of:
- a) SAL-S1 (SEQ ID NO:6);
  - 15 b) SAL-D4 (SEQ ID NO:8);
  - c) LpTK 2 (SEQ ID NO:10);
  - d) LpTK 3 (SEQ ID NO:12);
  - e) LpTK 4 (SEQ ID NO:14); and
  - f) LpTK 13 (SEQ ID NO:16).
  - 20 g) HpTK 5 (SEQ ID NO:24);
  - h) bpTK 1 (SEQ ID NO:25);
  - i) bpTK 2 (SEQ ID NO:26);
  - j) bpTK 3 (SEQ ID NO:27);
  - k) bpTK 4 (SEQ ID NO:28);
  - 25 l) bpTK 5 (SEQ ID NO:29); and
  - m) bpTK 7 (SEQ ID NO:30).

-17-

6. A homogeneous protein of human megakaryocytic origin which includes an amino acid sequence exhibiting sequence homology with the catalytic domain of tyrosine kinases of the c-kit family.
- 5 7. A homogeneous protein of Claim 6 in which the amino acid sequence is selected from the group consisting of:
  - a) SAL-S1 (SEQ ID NO:6);
  - b) SAL-D4 (SEQ ID NO:8);
  - 10 c) LpTK 2 (SEQ ID NO:10);
  - d) LpTK 3 (SEQ ID NO:12);
  - e) LpTk 4 (SEQ ID NO:14); and
  - f) LpTK 13 (SEQ ID NO:16).
  - g) HpTK 5 (SEQ ID NO:24);
  - 15 h) bpTK 1 (SEQ ID NO:25);
  - i) bpTK 2 (SEQ ID NO:26);
  - j) bpTK 3 (SEQ ID NO:27);
  - k) bpTK 4 (SEQ ID NO:28);
  - l) bpTK 5 (SEQ ID NO:29); and
  - 20 m) bpTK 7 (SEQ ID NO:30).
8. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FLT/FLK family of protein tyrosine kinases.
9. A protein of Claim 8 encoded by the nucleotide  
25 sequence (SEQ ID NO:5).
10. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO:6).

-18-

11. A protein of Claim 8 encoded by the nucleotide sequence (SEQ ID NO:17).
12. A protein of Claim 8 encoded by the amino acid sequence (SEQ ID NO: 18).
- 5 13. A protein of human megakaryocytic origin which exhibits significant sequence homology with the FGF receptor family of protein tyrosine kinases.
14. A protein of Claim 10 encoded by the nucleotide sequence (SEQ ID NO:7).
- 10 15. A protein of Claim 10 encoded by the amino acid sequence (SEQ ID NO:8).
16. A protein of human megakaryocytic origin which exhibits significant sequence homology with the NGF receptor family of protein tyrosine kinases.
- 15 17. A protein of Claim 14 encoded by the nucleotide sequence (SEQ ID NO:11).
18. A protein of Claim 14 encoded by the amino acid sequence (SEQ ID NO:12).

-19-

19. A DNA expression vector containing a DNA sequence of human megakaryocytic origin which hybridizes to a DNA fragment which hybridizes to DNA encoding an amino acid sequence present in the catalytic domain of a protein tyrosine kinase of the c-kit subgroup of protein kinases.
20. The DNA expression vector of Claim 17 containing a DNA sequence selected from the group consisting of:
- a) SAL-S1 (SEQ ID NO:5);
  - b) SAL-D4 (SEQ ID NO:7);
  - c) LpTK 2 (SEQ ID NO:9);
  - d) LpTK 3 (SEQ ID NO:11);
  - e) LpTk 4 (SEQ ID NO:13); and
  - f) LpTK 13 (SEQ ID NO:15).
  - g) LpTK 25 (SEQ ID NO: 22 and
  - h) HpTK 5 (SEQ ID NO: 23).
21. A cell transformed by the expression vector of Claim 17.

1/35

sals1 (160 bases)  
FLKI-LIKE

1	PTKI/3 PRIMERS	21	41
5' <u>ggatcctgtgcatcagtgacttagggctagggaacattctgtctcggaagcga</u> <u>cggtggt</u>			
		D P V H Q a L R A R N I L L S E S D V V	
61		81	101
gaagatctgtgactttggccttggcccggaacatctacaagagaccagctacgtccgcaa			
		K I C D F G L A R D I Y K D P S (Y) V R K	
121		141	PTKKW PRIMER
gcagcccggtgccctgaagtggatggcgccagaattc 3'			
		H A R L P L K W H A P E F	

FIGURE 1

2/35

sald4 (147 bases)-  
FGFR-LIKE

1	PTKI/3 PRIMERS	21	41
6'	<u>ggatccattcacagagacctagcagcacgcaacatcctgggtcttcagaggacctggtaacc</u>		
	G S I H R D L A A R N I L V S E D L V T		
61	81	101	
	<u>aaggtcagcgactttggcctggccaaagccgagcggaagggttagactcaagccggctg</u>		
	K V S D F G L A K A E R K G L D S S R L		
121	PTKKW PRIMER	141	
	<u>ccgtcaaatggatggctcccgaattc</u> 3'		
	P V K W H A P E F		

FIGURE 2

3/35

**LpTK2**

GTTGGAATTCCTTCGGGCGCCATCCATTTACCGGCAGCTTTATTTTCGTGTCTAGATTCA  
TAGATGTCTTCATTATCTACCTTAAAACTCTGGCAAGTCCAAAATCTGCTACTTTGTAG  
ATATTATGTTACCAACGAGGACATTCCT

**FIGURE 3A****LpTK3**

GTGCACAGGGATCTCGCGGCTCGGAACATCCTCGTCGGGGAAAAACCCCTCTCGAAAGTT  
GGGGACTTCGGGTTAGCCAGGCTTATCAAGGAGGACGTCTACCTCTCCCATGACCACAAT  
ATCCCCTACAAATGGATGGCCCCCTGAGGGAA

**FIGURE 3B****LpTK4**

G TTCACCGAGATCTCAAGTCCAACAACATTTTGCTGCTGCAGCCCATTGAGAGTGACGAC  
ATGGAGCACAAGACCCTGAAGATCACCGACTTTGGCCTGGCCCGAGAGTGGCACAAAACC  
ACACAAATGAGTGCCGC

**FIGURE 3C****LpTK13**

GTCAATCGTGACCTCGCCGCCCCGAAATGTGTTGCTAGTTACCCAACATTACGCCAAGATC  
AGTGATTTTCGGACTTTCCAAAGCACTGCGTGCTGATGAAACTACTACAAGGCCCAGACC  
CATGGAAAGTGGCCTGTCAAGTGGTACGCTCCGGAATGCATCAACTACTACAAGTTCTCC  
AGCAAAAGCGATGTCTGGTCCTTTGGAATTC

**FIGURE 3D**



1 TTCGAGCTCG CCGGACNTTG ATTATTGACT AGTATTATTAT AGTAATCAAT TACGGGGTCA TTAGTTTCATA GCCCATATAT GGAGTTCGCG GTTACATAAC  
AAGCTCGAGC GGCGTGAC TAATAACTGA TCANTAAATTA TCATTIAGTTA ATGCCCCAGT ATCAAGTAT CCGGTATATA CCTCANGGCG CAATGTATTG

101 TTACGGTAA TGGCCCGCTT GCGTGACGCG CCACGACCC CCGCCCATTTG ACCTCAATTA TCGAGTATGT TCCATAGTA ACCCAATAG GGACTTTCCA  
AATGCCATTT ACCGGGCGGA CCGACTGCGG GGTTCCTGGG GCGGGGTAC TCGAGTATTT ACTGCATACA AGGTATCAT TCGGTATATC CCTGANAAGT

201 TTGACGTCAA TGGGTGGAGT ATTTACGTA AACTGACCAC TTGGCAGTAC ATCAAGTGA TCATATGCCA AGTACCCCC CTATTGACGT CAATGACGGT  
AACTGCAGT ACCCACCTCA TAAATGCCAT TTGACCGGTG AACGTCATG TAGTTCACAT AGTATACGCT TCATGGGGG GATAACTGCA GTTACTGCCA

301 AATGGGCCG CCTGGCATTA TGCCCACTAC ATGACCTTAT GGCAGTTTCC TACTTGGCAG TACATCTACG TATTAGTCAAT CGCTATTACC ATGCTGATGC  
TTACCGGGC GACCGTAT ACGGGTCTAT TACTGGATA CCTGAAAGG ATGAACCTC ATGTAGATGC ATATATCATA CGGATAATGG TACCACATCG

401 GCTTTTGGCA GTACATCAAT GGGGTGGAT AGCGTTTGA CTCACGGGA TTTCACATC TCACCCCCAT TGACGTCAAT GGGAGTTTGT TTTGGCACCA  
CCAAACCGT CATGTAGTTA CCGGCACCTA TCGCCAAACT GAGTGCCCT AAGGTTTCA AGGTGGGTA ACTGCAGTTA CCTCAACA CAACCGTGT

501 AATCAACGG GACTTTCCAA ATGTGCTAA CAACTCCGCC CCATTGACGC AATGGGCG TAGGCGTGA CGGTGGAGG TCTATATAG CAGAGCTCGT  
TTAGTTGCC CTGAACGTT TTACAGCAT GTTGAGGCG GGTAACTGG TTACCCGCC ATCCGCACAT GCCACCTCC AGATATATTC GTCYCGACCA

601 TTAGTGAAC GTCAGATCC CTGGAGACGC CATCCACGCT GTTTTACCT CCTAGAGA CACCGGACC GATCCAGCT CCGGGCCCG GAACGCTGCA  
AATCACTTG CAGCTAGCG GACCTCTGCG GTAGTGCGA CAAACTGGA GGATCTTCT GTGGCCCTGG CTAGGTGCA GCGCCCGCC CTGCCACGT

FIGURE 4A

5/35

701 TTGGAACGGC GATTCGCCGT GCCAAGAGTG ACGPAGTAC CGCCTATAGA GTCTATAGGC CCALITGGCT TCGTTAGAAC GCGGCTACAA TTATATACATA  
AACCTTGGCC CTAAGGGGCA CGGTCTCACC TGCATTTCATG GCGGATATCT CAGATATCCG GGTGAACCGA AGCANCTTIG CCGCGATGTT ANTIATGTAT

801 ACCTTATGTA TCATACACAT ACGATTTAGG TGACACTATA GAATAACATC CACTTTGCCCT TTCTCTCCAC AGGTGTCAC TCCCAGGTCC AACTGCACCT  
TGGATATCAT AGTATGTGTA TGCTAAATCC ACTGTGATAT CTATTTGTAG GTGAACCGA AAGAGAGGTG TCCACAGGTG AGGTCCAGG TTGACGTGGA

901 CGGTCTATC GATTGAATTC CCGGGGATC CTCFAGAGAT CCCTCGACCT CGAGATCCAT TGTGCTGGG CGATTTCTTT ATCACTGATA AGTTGGTGA  
GCCAAGTAG CTAACCTTAAG GGGCCCTAG GAGATCTTA GAGAGCTGGA GCTCTAGGA ACACGACCCG CCTAAGAA TACTGACTAT TCACCCACCT

1001 CATATTATGT TTATCAGTGA TAAAGTGTA ACGATGACAA AGTGCAGCC GAATACAGTG ATCCGTGCCG CCTAGACCT GTTGAACGAG GTCGGCGTAG  
GTATAATACA ATATGCACT ATTTCACAGT TCGTACTGTT TCAACGCTGG CTATGTGAC TAGGCACGGC GGCATCTGGA CAACTTGCTC CAGCCGATC

1101 ACGGTCTGAC GACAGGCNAA CTGGCGGAA CTTGGGGGT TCAGAGCCG GCGTTTACT GGCATTTCAG GAACAGCGG GCGTGTCTG ACGCACTGGC  
TGCCAGACTG CTGTGCGTTT GACCGCCTTG CCAACCCCA AGTCGTGGC CCGCAATGA CCTGAAGTC CTGTTCGCC CCGACCGAGC TGGGTGACCG

1201 CGAAGCCATG CTGGCGGAGA ATCATAGCAC TTGGGTGCCG AGAGCCGACG ACGACTGGCG CTCATTTCTG ACTGGGAATG CCGCGAGCTT CAGCAGGCG  
GCTTCGGTAC GACCGCCTCT TAGTATCTG AAGCCACGGC TCTCGGCTGC TCGTACCCG CAGTAAAGAC TGACCTTAC GGCCTGGA GTCCCTCCG

1301 CTGCTGCCCT ACCGCCAGCA CAATGGATCT CGAGGGATCT TCCATACCTA CCAGTTCTGC GCTGTGAGGT CGCGGCCGCA CTACTCTTG ATGATTTACT  
GACGAGCGGA TGGCGGTCTG GTTACCTAGA CTTCCCTAGA AGGTATGAT GGTCAAGACG CCGACCTCCA GCGCCGCGCT GATGAGAAC TACATATGA

FIGURE 4B

1401 CATATTACCA AGGAATAACT GCGGGGCACA GGGTCAGGTG CTGAAGGGAC ATTGTGAGAA GTGACCTAGA AGGCAAGAGG TGAGCCCTCT GTACCGCTGG  
 GTATANTGGT TCCTTATTGA CCGGCCGTGT CCGAGTCCAC GACTTCCTG TAAACACTTT CACTGGATCT TCCGTCTCC ACTCGGAGA CAGTCCGACC

1501 CATAAAGGGCC GCTTGAGGGC TCTTTGGTCA AGCAGTAACG CCAGTCTCTG GUAGGCCACC TGTTACTCAG CAGACCATGA AAGGGCTCT CCCTTTCTT  
 GTATTCCTGG CCAACTCCCG AGAACCAGT TCGTCAATGC GGTACAGAC CTTCCGTGG ACAATGAGTC GTCTGTACT TTCCCGAGA GGGAAAGGA

1601 GCAGCAGTCA GGGAACTC TGCTCCACCA GCTTCFTGTG GGAGGCTGGA TATTATCCAG GCTGCCCCG AGTCATCCG AGCCCTAAC CCTCCCTGTG  
 CCTCCTCAGT CCTTGTGAG ACGAGGTGGT CCAAGACAC CCTCCGACT ATANTAGGTC CGACGGGCG TCAGTAGGCC TCCGATTTG GAGGGACAC

1701 GTGCTTCACT GGTACACTC CTGTGCTCACT TCAATGCTCC TCTTGGCTC CTGTTCTCTG TTGGAGTTT GTATAGATA GCAGAGAA TAGCGAAGT  
 CACGAAGTCA CCAGTGTGAG GAACAGGTGA AAGTAGGAG AGAACGGAG CACCAAGGAG AACCTTCAA CATCATCTAT CGTCTCTT ATGCTTTCA

1801 CTTAAGTCT TTGATCTTC TTATAAGTGC AGAGAGAA TGTGAGCTA TGTGCTTC TCTCTCTG CTTCAGCTAC CTGAAGCCG TTTCTTGTCT  
 GAATTCAGA AACTAGAAAG AATATTACG TCTCTCTT ACAGTGCAT ACGACGGAG AGAGAGAGC GAAGTGGATG GACTTCGGG AAAGAAAGA  
 349 O R P G S E O R

1901 ATACCTGCTC TCTATCTGCT CACACTCCTC CGAGGCCAGC ACCATCCAC TGTCTGTCTG GTGTCCACA GAGCTTTCT AGTCTGTTG GGTCTATGGG  
 TATGACGAG AGATAGACGA GTGTGAGGAG GTCTCGGTG TGTAGGTG ACAGACAGAC CACAGGTGT CTCGGAACA TCCAGCAACC CCAGTACCC  
 341 Y R S E I Q E C E E S A L V H G S D Z Q N D V S G K Y T T P T H P

2001 AATTCCTCAA ATGCTTCTAT CTTGGAGGAA CCAAGGGTCT CAGCCCTCT GGCAGGCAC CCGGGAAGG ACACCAATT GTATACCTG CCGCCACGC  
 TTAAGGAGTT TACAGAAGTA GACTCTCTT GTTCCCGA GTCCGGGAGA CUGTCTCTG GGCCTTTCC TGTGGGTCAA CATATGGAC CGCCGTGCG  
 308 P E Z F T K H R S S G R T E A G R A L C G P F S V W N Y Y R A A L

FIGURE 4C

2101 TGTGGCGCTG CAGGCTTGGC GGGCTGTCTT CAGCCTCAGC CTGGGCGATG TGTAGGGCCA TGGTGGACAC CTGGGAGAG CTGCCCTCTT CTGAGCTCTG  
 ACACCGCGAC GTCCGACCG CCCGACGCA GTCCGACTCG GACCGCTAC ACATCCCGT ACCACCTTG GACGCTCTC GACGGAGAA GACTCGAGAC  
 215 S H R Q L S P P S D E A D A Q A I H L A M T S V Q S F S G E Z S S Q

2201 AGAGCTGGCG GGGGCAATCC AGACTCTCTC TCTCTCTTG AGGCCCTGC CCTGGAGAG GTCCCCAGG ATCTCCACCA GCTCCGAGAA TGCAGGTCTC  
 TCTCGACCG CCCGCTACG TCTGGAGGAG AAGGAGNCC TCCGGGACG GGACCTCTC CAGGGCTCC TAGAGTGT CGAGGCTCTT ACGTCCAGAC  
 241 S S R P A M C V E E E Q L G R G O L L D G L I E V L E S F A P R

2301 GCCTTGGGT CTCCGGACCA GCAATTCAGC ATATGGGCG ATATGGGCGG TCCGGGCGG TCATCTCTT GCCCTCTCTC AGCCGCTGGC  
 CGGAACCCCA GAGCCTGTG CTCAAGTCTG TACTACGCG CATACGCCC TCACCGTCTG AGGCCCGG AGTAGGAACTA CGGCAGAGAG TCGGCGACCG  
 208 A K P D G S W C N L M I R R I A P T A L E P A R M R T G D R L R Q

2401 AGACTCTCTC ATGATCTGC AGCCAGGGT ACGGGAGGC CCCAGAGAG AAGTCTCTC AGAGAGCAC CCCAAGGAC CACACGTCTC TCTGGGTGGT  
 TCTGAGGAG TACTAGAC TGGGTCCCA TGCCTCTCG Y P S A G L S P I E W L L V G P S W V D S Q T T

175 C F E E N I Q V G P Y P S A G L S P I E W L L V G P S W V D S Q T T

2501 GTACACCTTG TCGAGATGC TTTCAGGGC CATCCACTTC AGGGCAGCC GGGCACTGCC CTTCGGGAG TACTCGGGT CTTCGTAGAT GTCCCGGCA  
 CATGTGGAC AGCTCTACG AAGTCCCG GTAGGTGAG TCCCGTCCG CCCGTGACG GAACGCTGC ATCAGCCCA D K Y I D R A  
 141 Y V K D F Y S E P A M W K L P L R A S G K R V Y D P

2601 AGGCCAAGT CACAGATCTT CACCAGTCT CTTTCGACA GCAGATGTT CCGAGCAGCC AGGCTCTGT GGATGCACTT TCGGAGGCC AGGAATCCA  
 TCCGCTTCA GTCTTAGA GTGGTGCACC GAAGGCTGT CGCTTACAA GGCTCTCGG TCCAGAGCA CCTAGTGAA AGCCCTCTCG TCCTTGAGGT  
 108 L G F D C I K V V D S E S L L I N R A A L D R H I C K R S A L P E

2701 TCCCTCTGCG CACCTGGAG CTGTAGCAGA CAAGATCTT CATGTCAGC GGGCTCAGC ACAGTCTCT AGCTCTTGG TCTGGAGAA CCCGCTCTCG  
 AGGAGACCG GTGGACCTT GACATCTCT GTTCTAGAG GTACCACTCG CCCGACTCG TCTCAGGAG TCGAGAAC AGACTCTCTT GGGCGGAGCG  
 75 M G R A V Q F S Y C V L D E M T L P S L W L D E A E Q D P S A R R A

FIGURE 4D

FIGURE 4E

2801' TCCGCCCTCG GTCTTCGAGA ACCGGCGGAA GAGGACCCYG TGGTGTCTCC CCGCGCGCT CCGATCCAGC CTGGCGAGCT CCACCATGGC GCGGAAGCGT  
 AGCGGGAGC CAGAAGCTCT TGGCGGCTT CTCTGGGAC AGCAGCGAGG GCGCGCGGA GGCTAGGTGG GACCGCTGA GTGTATCCG CCGCTTCGCA  
 .41' G G E T K S P R A F L V R D S S G P R R R D L R A L E V M A R F R  
 2901 CCGCGTGTCT CCGGACACTT CTCTGCGGA TCCACGAAGC TGGCTCGAGG GCGCCAGTC GTCCGCCGA GAGCGGCTC CATTCCCCG CCGCGCGCGG  
 GCGCGGACGA GCCCTCTGAA GAGGACGCTT ACCTGTCTCG ACCGAGCTCC CCGCGGTGAG CAGCGGCTC GTCCCGGAG GTACGGGGC GTACGGGGC  
 B G R Q E P S K E  
 3001 CCGCGCGCAG CCGCGCGCT' CACCGCGCAG GGGCTCGGC CCGACTCTA GAGTCGACCT GCAGAGCTT GCGCGGCTG GCGGAGCTG TTTATTCAG  
 GCGGGGGTC CCGCGGCGA GTGGCTCTC CCGGAGCGC GCGCTGAGT CTCAGCTGA CCGCTTCGA CCGCGGTAC CCGGTTCAG AATACGTC  
 3101 CTTATAATGG TTACAAATAA AGCAATAGCA TCACAAATT CACAAATAA GCATTTTTT CACTGCNTC TAGTGTCTT TTGTCCAAAC TCATCAATCT  
 GAATATTACC AATGTTATT TCGTTATCT AGTGTATAA GTCTTATT CTAATAAAA GTACGTAAG ATCAACCA ACACGTTT AGTAGTACA  
 3201 ATCTATCAT GTCTGGATCG ATCGGGAATT AATCGGCGC AGCACCATGG CCGCAATAA CCGCTGAAAG AGCACTTG TTAGGTACCT TCTGAGCGG  
 TAGAATAGTA CAGACCTAGC TAGCCCTTAA TTAGCGCGC TCGTGGTACC GGACTTTAT CCGACTTTC TCCTTGAACC AATCCATGGA AGACTCGGC  
 3301' AAAGAACCAG CTGTGGAATG TGTGTCACTT AGGTGTGGA AATGCCCCAG GCTCCCCAGC AGCAGAGCT ATGCAAGCA TGCATCTCA TTAGTCAGCA  
 TTCTTGGTC GACACCTTAC ACACAGTCAA TCCACACCT TTCAGGGGTC CAGGGGTG TCGTCTTCA TACGTTCTT ACGTAGAGT AATCAGTCT  
 3401' ACCAGGTGTG GAAAGTCCC AGGCTCCCC GCAGCGAGAA GTATGCAAG CATCATCTC AATTAGTCAG CAACCTACT CCGCGCCCTA ACTCGGCCCC  
 TGTCCACAC CTTICAGGG TCCGAGGGT CGTCTGCTT CATAGTTTC GTACGTAGAG TTATCAGTC GTTGGTATCA GCGCGGGAT TCAGCGCGGT

9/35

FIGURE 4F

3501 TCUCGUCCTT AACTCGGCC AGTCCGGCC ATTCTCGCC CCATGGCTGA CTATTTT TTTATATGC AGAGGCCGAG GCGCCTCGG CCTCTGAGCT  
 AGGGCGGGA TTGAGCGGG TCAGGCGGG TAAAGCGGG GGTACCGACT GATTAAAAA AATAAATACG TCTCCGGCTC GCGCGGAGCC GGAGACTCGA

3601 ATTCCAGAG TAGTGAGAG GCTTTTGG AGGCTAGGC TTTTGCMAA AGCTGTAAAC AGCTTGGCAC TGGCGGTGT TTTACAAAGT CGTGACTGGG  
 TAAGGTCTC ATCACTCTC CGAARAAACC TCCGATCCG AAACGTTTT TCGACAAATG TCGAACCGTG ACCCGACGA AATGTGCA GCACTGACCC

3701 AAACCCCTGG CGTACCCAA CTTAATCGC TTGCAGCACA TCCCCCTTC GCACTGTGC GTAATAGCGA AGAGGCCCG ACCGATCGCC CTTCCTCAACA  
 TTTTGGGACC GCAATGGGT GANTAGCG AACGTCTGT AGGGGGGAG CGGTGACCG CATATCGCT TCTCCGGCG TGGTAGCGG GAAGGTTGT

3801 GTTGGGTAG CTGAATGGG ATGCGGCTT GATGGGTAT TTCTCTTA CGATCTGT CCGTATTCA CACCGCATAC GTCAAGCAA CCATAGTACG  
 CAACCATCG GACTTACCG TTACCGCGA CTACGCCATA AAAGAGGAT GGTAGACAC GCATTAAGT GTGGGTATG CAGTTCTGT GGTATCATCG

3901 CGCCCTGTAG CGCGCATTA AGCGGCGCG GTGTGTGCT TACGCGCAG GTACCGCTA CACTTGGCAG CGCCCTAGCG CCGCTCTT TCGTTTCTT  
 GCGGCACATC GCGCGTAT TCGCGCGCC CACACCACA ATGCGGTCTG CACTGGGAT GTGAACGGTC GCGGGATCG GCGCGAGGA AGGNAAGAA

4001 CCTTCTCTT CTGCGCAGT TCGCGGCTT TCCCGTCAA GCTCPAATC GGGGCTCC TTATGGGTG CGATTTAGT GTTACGCA CCTCGACCC  
 CGGAAGGAA GAGCGGTGA AGCGCGGA AGGGCACT CGAGATTAG CCGCGAGGG AATCCCAAG CTAATATCAC GAATGCCGT GGAGCTGGG

4101 AAAAAGCTT ATTTGGCTA TGGTCACT AGTGGGCAAT CGCCCTGATA GACGTTTTT CCGCTTTGA CGTTGGAGT CAGTTCTT AATAGTGGAC  
 TTTTGTGAC TAAACCCACT ACCAAGTGA TCACCGGTA GCGGACTAT CTGCAAAA GCGGGAAT GCAACCTAG GTCAAGAAA TTATCACCTG

10/35

FIGURE 4G

4201 TCTTGTCTCCA AACTGGGAACA ACACTCAACC CTATCTCCGG CTATCTCTTT GATTIATTAAG GGATTTTGGC GATTTCGGCC TATGTGGTAA AANAATGAGCT  
AGAACAAAGGT TTGACCTTGT TGTGAGITGG GATAGAGCCC GATAGAAAA CTAATATTC CTAAGCCCG CTAAGCCCG ATAACCAAT TTTTACTCGA

4301 GATTTAACAA AATTTAAACG CGAATTTTAA CAAATATTA ACGTTTACAA TTTATGGTG CACTCTCAGT ACAATCTGCT CTGATGCCG ATAGTTAAGC  
CTAAATTTGT TTTAATTTGC GCTTAAATTT GTTTTATAT TGCATATCTT AANATACCAC GTGAGAGTCA TGTTAGACCA GACTACGGCC TATCAATTCG

4401 CAACCTCCCT ATCGCTACGT GACTGGGTCA TGGCTGGCC CGACACCCG CCAACACCCG CTGACCCGCC CTACAGGGGT TGTCTGTCTC CGCATCCG  
GTTGAGGCGA TNGCGATGCA CTGACCCCACT ACCGACCCG GGTCTGGCC GACTGGCCG GACTGGCCG ACAGACGAG GCCGTAGCG

4501 TTACAGACAA GCTGTGACCG TCTCCGGGAG CTGCATGTGT CAGAGTTTT CACCGTCATC ACCGAAACG CGGAGGAGT ATTCTTGAG AGCAAGGCG  
AATCTCTGT CGACACTGGC AGAGGCCCTC GACGTACACA GTCTCCAAA GTGGCAGTAG TGGCTTGGC CGCTCCGTC TAAGNACTTC TGTCTTCCG

4601 CTGCTGATAC GCTATTTTT ATAGCTTAA GTCATGATAA TAATGGTTTC TTACAGCTCA GGTGGCACTT TTGGGGGAAA TGTGGCGGA ACCCTATTT  
GAGCACTATG CGCATAAAA TATCCANTTA CAGTACTATT ATTACCAAG ATCTGCAGT CCACGTGA AGCCCCCTT ACACCGGCT TGGCGATAA

4701 GTTATTTTT CTAAATACAT TCAATATGT ATCCGCTCAT GAGACAAZAA CCTGATAA TGCTTCATA ATATTGAAA AGGAAGATA TGAGTATCA  
CAAAZAAAA GATTATGTA AGTTATACA TAGGCGAGTA CTCTGTATT GGGACTATTT ACCAAGTTAT TATAACTTTT TCTTCTCAT ACTCATAGT

4801 ACATTTCCCT GTCCGCCCTTA TTCCCTTTTT TCGGGCATTT TCCCTTCTG TTTTGTCTA CCCAGAACG CTGGGAAAG TAAAGATGC TGAAGATCAG  
TGTAAAGCA CAGCGGGAAT AAGGAAAAA ACCCGTAA TCGCGTAA GGGCTTTGC GACCACTTC ATTTCTACG ACTTCTAGC

FIGURE 4H

4901 TTGGGTGCAC GAGTGGGTTA CATTGAACTG GATCTCAACA GCGGTAAAGAT CATTGAGAGT TTTCGCCCCG AAGAACGTTT TCCANTGATG AGCATTFTT  
 AACCCACGCG CTCACCCCAAT GTACCTTGAC CTAGAGTTGT CCCCATTCTA GGAACCTCA AAGCGGGG TTCTGCANA AGGTACTAC TCGTGAANA

5001 AAGTCTCTCT ATGTGGCGCG GTATTATCCC GTGATGACCG CGGGCAAGAG CAACCTGGTC GCCGATACA CTATCTCAG AATGACTGG TTGAGTACTC  
 TTCAAAGACA TACACCGCGC CATATAAGG CACTACTCG CCCCCTCTC GTTGAGCCAG CGGCTATGT GATAAGTIC TTAAGTGAAC AACTCATGAG

5101 ACCAGTCACA GAAAGGCATC TTACGGATGG CATGACAGTA AGAGATTAT GCAGTCTGCG CATACCATG AGTATACA CTGCGGCCA CTTACTTCTG  
 TGGTCACTGT CTTTTCGTAG AATGCTTACC GTACTGTCT TCTCTTATA CGTCACGAG GTATTGGTAC TCACTATTGT GACGCGGTT GAATGAAGAC

5201 ACAACGATCG GAGGACCGAA CGAGCTAACC GCTTTTTCG ACACATGGG GGATCATGTA ACTGCGCTG ATCGTTGGA ACAGGACTG AATGAAGCCA  
 TGTTCGTAGC CTCCTGGCTT CCTCGATTGG CGAAAAAACC TGTGTATACC CTTAGTACAT TGACGGGAAC TAGCAACCTT TGGCTCGAC TTAATTCTGT

5301 TACCAACGA CGAGCGTGAC ACCACGATGC CAGCAGCAAT GGCACCAAC TTGCGCAAC TATTAACTGG CGAATCTATT ACTCTAGCTT CCGCGCAACA  
 ATGTTTGTCT GCTCGCACTG TGGTGCTACG GTCTGTCTTA CCGTGTGTC CCGTGTGTC AATATGACC GTTGTATGNA TGAGATGNA GGGCGTTGT

5401 ATTAATAGAC TGGATGGAGC CGGATAAAGT TGCAGACCA CTCTGCGCT CGGCTCTTC GGTGGCTGG TTATGCTG ATAAATCTGG AGCGGTGAG  
 TAATATCTG ACCTACCTCC GCCTATTCTA ACCTCTCTGT GAAGCGGA CCGCGGAGG CCGACCGACC AATATAGAC TATTAGACC TCGCGCACTC

5501 CGTGGGTCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAAGCCCTC CCGTATCGTA GTTATCTACA CGACGGGAG TCAGGCACT ATGATGAAC  
 GCACCCAGG CCCCATAGTA ACCTCTCTAC CCGGTCTAC CATTCGGGAG GGCATAGCAT CATTAGTGT GCTGCCCTC ACTCCGTGA TACTACTTG



FIGURE 4I

5601 GAAATGACGA GATCGCTGAG ATAGGTGCCT CACTGATTAA GCATTGGTAA CTGTGAGACC AGCTTACTC ATATATACTT TAGATTGATT TAAACTTT  
CTTTATCTGT CTAGCGACTC TATCCACGGA GTGACTAATT CGTAACCATI GACAGTCTGG TTCAANTGAG TATATATGAA ATCTAACTAA ATTTTGAAK

5701 TTTTAAATT AAAGGATCT AGGTGAAGAT CCTTTTIGAT AATCTCATGA CCAAATCC TAAAGTGAG TTTTCGTTC ACTGAGGTC AGACCCCGTA  
AAANTTAA TTTTCTAGA TCCACTTCTA GGAAACTA TTAGAGTACT GGTTTAGG AATTGCATC AATAGCAAG TGACTGCCAG TCTGGGCAT

5801 GAAAGATCA AAGATCTTC TTGAGATCTT TTTTCTGC GCGTATCTG CTGCTTGCA ACAAATAAC CACCGCTACC ACCGTGCTT TGTTCGCGG  
CTTTCTAGT TTCTAGAG AACTCTAGGA AAAAAAGAG CGCATTAGAC GACGATGTT TGTTTTGTG GTGCGATG TCGCCACCAA ACNAACGGCC

5901 ATCAAGAGCT ACCAATCTT TTTCGAGG TAACTGGCT CAGCAGAGG CAGATACCA ATACTGTCT TCTAGTGTAG CCGTAGTTAG GCCACCACTT  
TAGTCTCGA TCGTTGAG AAAGGCTTC ATTGACCGAA GTCTGTCTCC GTCTATGTT TATCAGGA AGTACATC GCATCATC CCGTGTGAA

6001 CAAGAACTCT GTAGCACCGC CTACATACCT CGCTCTGCTA ATCTGTATC CAGTGGCTGC TGGCAGTGGC GATAGTCTT GTCTTACCG GTTGGACTCA  
GTTCTTGAG CATCGTGGC GATCTATGGA GCGAGACGAT TAGGACATC GTCCACCGAG CCGTCAACC CTATTCAGCA CAGATGGCC CAACCTGAT

6101 AGACGATAGT TACCGATTA GCGGCAGCG TCGGGCTGAA CCGGGGTTTC GTGCACACAG CCCAGCTTGG ACCGACGAC CTACACCGNA CTGAGATACC  
TCTGCTATCA ATGCGCTATT CCGGTCGCC AGCCCGACTT GCGCCCAAG CACGTGTCTG TCGCTGTCTG GATGTGGCTT GACTCTATGG

6201 TACAGCGTGA GCATTGAGAA AGCGCCAGC TTCCCGAAG GAGAAAGCG GACAGGTATC CGGTAAGCG CAGGTCTGGA ACAGGAGAGC GCACGAGCGP  
ATGTCGCACP CGTAACCTCT TCGCGGTGCG AAGGCTTCC CTCTTTCCG CTGTCCATAG GCAATTCGCC GTCCAGCCT TGTCTCTCTG CGTGTCTCC

13/35

FIGURE 4J

6301 GCTTCCAGGG GGAACGCCT GGTATCTTAA TAGTCCTGTC GGGTTTCGCC ACCTGTGACT TGAGCTGCGA TTTTGTGAT GTCGTCAGG GGGCGGAGC  
CGNAGGTCCC CCTTTGCCGA CCATAGAAAT ATCAGGACAG CCCAAGCGG TGCAGACTGA ACTGCGAGCT AAAACACTA CGAGCAGTCC CCCCCTCTG

6401 CTATCGAAA ACGCAGCAA CCCGGCCTTT TTACGTTTC TGGCCTTTTG CTGGCCTTTT GTCACATCT TCTTCTCTCC GTTATCCCTT GATTCCTG  
GATACCTTT TCGGTCGTT GCGCCGAAA ATGCCAAG ACCGGAAC GACCGAATA CCACTGTACA AGAAGGACG CAATAGGGA CTACACACC

6501 ATAACCTAT TACCGCCTTT GATGAGCTG ATACCGCTCG CCGCAGCGA ACGACGAGC AGTGACGAG GAACGGAG AGCGCCCAAT  
TATTGGCATA ATGCGGAAA CTCATCGAC TATGCGAGC GCGTGGCT TGTGGCTCG CGTGGCTCAG TCACCTGCTC CTTGGCCTC TCGCGGTTA

6601 ACGCAACCG CCTCTCCCG CGCGTTGGC GATTCATTAA TCCAGCTGC ACGACAGTT TCCGACTCG AAGCGGCA GTGAGCGCA CGCAATTAAT  
TGGTTTGGC GGAGAGGGC GCGCAACCG CTAGTAAT AGGTGACCG TGCTGTCCA AGGCTGACC TTTCGCCCT CACTCGCTT CGCTTAATTA

6701 GTGAGTACC TCACTCATTA GGCACCCAG GCTTACACT TTATGCTTC GCTGTATG TTGTGTGGA TTGTGAGCG ATAACTAT CACACAGGA  
CACTCATGG AGTGATAT CCGTGGGTC CGAATGTGA AATACGAGG CCGAGCATAC AACACCTT AACACTGCC TATTGTAAA GTGTCTCTT

6801 ACAGCTATGA CCATGATTAC GAATTAA  
TGTGATACT GGTACTAATG CTTAATT

FIGURE 5A

```

1  TTCCAGCTCG CCGGACATTC AATTATTCACCT AGTAAATCAAT TAGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC
   AACCTCGAGC GGGCTGTAACT TAATAAACCTA TCAATAAATA TCAATTAATTA ATGCCCCAGT ATGCAAGTAT CCGGTATATA CCTCAGGGCG CATTGTATTG

101 TTACGGTAA TTGCCCCGCTT GGCTGACCGC CCACGACCC CCGCCCATTC AGGTCAATAA TGACGTATGT TCCCATAGTA ACGCCAATAG GGACTTTCCA
   AATGCCATTT ACCGGGCGGA CCGACTGGCG GGTGTCTGGG GCGGGTAACT TGCAGTTATT ACTGCATACA AGGTATTCAT TCGCGTTATC CCTGAAAGGT

201 TTACAGTCAA TGGGTGGAGT AATTACGCTA AACTGCCCCA TTGGCAGTAC ATCAAGTCTA TCATATGCCA AGTAGGCCC CTATTGACGT CAATGACGGT
   AACTGCAGTT ACCCACCCTA TAAATGCCAT TTGACGGGTG AACCGTCAAT TAGTTACAT AGTATACGGT TCATGCGGG GATAACTGCA GTTACTGCCA

301 AAATGGCCCG CCTGGCATTA TGGCCAGTAC ATGACCTTAT GGGACTTTCC TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC
   TTTTACCGGC GGACCGTAAT ACGGTTCATG TACTGGATTA CCTGAAAGG ATGMACCGTC ATGTAGATGC ATATCAGTA GCGATAATCG TACCACCTAGC

401 GGTTTTGGCA GTACATCAAT GGGCGTGGAT AGCGTTTGA CTCACGGGA TTTCGAAGTC TCCACCCCAT TGACGTCAAT GGGAGTTTGT TTTGGCACCA
   CCMAAACCGT CATGTAGTTA CCGGCACCTA TCGCCMAACT GAGTGCCCC CTAGGTTTCA AGGTGGGTA ACTGCAGTTA CCTTCNAACA AAACCGTGGT

501 AAATCAACGG GACTTTCCAA AATGTCTAA CAACTCCGCC CCATTGACGC AAATGGGCG TAGGCGTCTA CCGTGGAGG TCTATATAAG CAGAGCTCGT
   TTTAGTTGCC CTGAAGGTT TTACAGCAAT GTTGAGGCG GTTAACCTGC TTACCCGCC ATCCGCACAT GCCACCTCC AGATATATTC GTCTCGAGCA

601 TTAGTGAACC GTGAGATCGC CTGGAGACGC CATCCACGCT GTTTTGACCT CCATAGMAGA CACCGGGAC GATCCAGCCT CCGCGGCCCG GAACGGTGCA
   AATCACTTGG CAGTCTAGCG GACCTCTGCG GTAGGTGCGA CMAAAGTGA GGTATCTTCT GTGGCCCTCG CTAGGTGCGA GCGCGCCGCC CTTCGCCCGT

701 TTGGMAACCG GATTTCCCGT GCCAAGAGTG ACGTAAGTAC CGCCTATAGA GTCTATAGGC CCACCTGGCT TCGTTAGAAC GCGGCTACAA TTAATACATA
   AACCTTGGCG CTAGGGGCA CCGTCTCTAC TGCATTCTATG GCGGATATCT CAGATATCCG GGTGAACCGA AGCAATCTTG CGCCGATGTT AATTATGTAT

```

15/35

FIGURE 5B

801 ACCTTATGTA TCATACACAT ACGATTTAGG TGACACTATA GAATMACATC CACTTTGCCT TTCTCTCCAC AGGTGTCCAC TCCAGGTCC AACTGCACCT  
TGGAAATACAT AGTATGTGTA TGCTAAATCC ACTGTGATAT CTATTGTAG GTGNAACGGA AMGAGAGGTG TCCACAGGTG AGGTCCAGG TTGACGTGGA

901 CGGTTCTATC GATTGAATTC CCCGGGATC CTCTAGAGAT CCCTCGACCT CGAGTCGACT TTTTTTTTTT TTTTGTAGG CCAAAGGGTA CTCTTTTTTC  
GCCAAGATAG CTAACCTAAG GGGCCCCCTAG GAGATCTCTA GGGAGCTGGA GCTCAGCTGA AAAAAAANA AAAAACATCC GGTTCCTCAT GAAGNAANAAG

1001 TTTATTAAAT ACTCAGAGT CTAGGCCACA GCAATCTACT GTTCTCTCTCT CATTTTCTTA AACTATTTTG ATACCTATTT CTCAGACTTT ATGGGTATTT  
AAATAATTA TGAGTCTTCA GATCCGTGT CGTTAGATGA CAAGAGGAGA GTAAAGGAT TTGATAAAAC TATGGATAAA GAGTCTGANA TACCCGATNA

1101 AGACATTTCT CACATTTCCA TAGATAATAA CTCATCCGTT TTGCMACCCTG ATTCTCAATA TTAAGAGATT AAACTAATG TATATGACTC TCAGTTGACA  
TCTGTAAAGA GTGTAAAGGT ATCTATTATT GAGTAGGCM AACGTTGGAC TAAGAGTTAT AATTCTCTAA TTTTGATTAC ATATACTGAG AGTCMACTGT

1201 CATACCTGAG TACAGAAAA TTCCATCATT TCCTTCTGCA AATGAAAAA GACTTGGTTT TCTCAACAGC TGCATCATTT TTTTATGCAT AGAAAAAAT  
GTATGACTTC ATGTCTTTT AAGGTAGTAA AGGAGACGT TTTACTTTTT CTGAAGCANA AGAGTTGCG ACGTAGTAAA AAAATACGTA TCTTTTTTTA

1301 GTGCAATTAC TCCAAGTACA ATCAAGTCAT TTAACATGGC TTTACCATCA TTGTAGTTAC AGGATATTTT AAAAGAGAAA AAAAAATCTC AAAGCACAGG  
CACGTTAATG AGGTTTCATGT TAGTTTCAGTA AATTGTACCG AANTGGTAGT AACATCAATG TCCATATAAA TTTTCTCTTT TTTTGTAGG TTTCTGTCTC

1401 TCCTGCTGTG CAGCAAGCA ATCAAAATCC TTCAATATAA CAGCCTGATG GGATTCAGCA ATCTGAGGA TAATGAATA CCACCTTAAT CAGTAAACAG  
AGGACGACAC GTCGTTTCTGT TAGTTTAAGG AAGTATTATT GTCGGACTAC CCTAAGTCGT TAGACTCCTT ATTACTTAT ATTACTTAT GGTGAGATTA GTCAATTGTC

1501 GAAATGCTA CAACAGTCAC TGAGTAAAA TTGCACTATC ATCTGTTGAT TCTCTGATC GACATTCAA ACAATAATG GAAATGTAAG TATCTCTTAA  
CTTTTACGAT GTTGTCACTG ACTCATTTTT AACCTGATAG TAGAGAACTA AGAGAACTAG CTGTAAAGTT TGTATTATC CTTTACATTC ATAGACAAAT

16/35

FIGURE 5C

1601 AAAGAAAT AACTGGTTT AGTGTGCTTA ATTTTACCAG GCAGTGAGGA AATTATATAT CACCTTGAAT GTCTGCAGT GTTGCCCAAT CAATAAATG  
 TTTCTTTTAA TTGNAACCAA TCACACGAAT TAAATGGTC CGTCACTCCT TTAATATATA GTGGAAGTGA CAGGACGTCA CAACGGGTCA GTTATTTTAC

1701 CACAAATAT CTTTTTCATA ATACATGGCC AACTTTATCC TATCACTTGA ATATGTCAGG ATAACTGAT TGTGCAGTTG GTTGATAACA TTGTATTTTG  
 GTGTTTATTA GAATAAGTAT TATGTACCGG TTGAATAAGG ATAGTGAAT TATACAGTCC TATTTGACTA ACACGTCAC CAACTATTGT AACATNAAC

1801 GAATGATTA TTGAATTTG TTTTGCTACT TTATTAATTTG ATATCTCTCT CCAGTGTTCA TCTTATGAAG TTATTTGCAT CTGAATATGA AGAGTCTGTT  
 CTTACCTAAT AACTTAAAC AAACGATGA AATAATNAAC TATAAGAAGA GGTCAAAAGT AGAATACTTC AATAAACGTA GACTTATACT TCTCAGACAA  
 506 O R I F N N A D S Y S S D T

1901 TCAAAATAGT CTTCAAGTTT CCAACGCAGT GTCTCAATG TAGTGTGTTT CTTAGGCTCT GCATTCAGC ACTCCAACAT GATGTTGTAA AATTGCTGTG  
 AGTTTATCA GAAGTCAAA GGTGCGTCA CAGAGTTTAC ATCCAGCAAG GAAATCCGAGA CGTAAGGTG TGAGGTGTGA CTACAACATT TTAACGACAC  
 492 E F Y D E L K W R L T E F T P R E K P E A N W C E L M I N Y F Q Q

2001 GACAGTTGGA TGGTTCGGA AGTCTATAGT TTTGAGCCAA CATCTGGATT ACCTGGGCAC CTGTCAATACC ACTGTAAGC ATTTGCCCAT AAGTAATGAT  
 CTGTCAACCT ACCAAGCCT TCAGATATCA AACTCGGT GTAGACCTAA TGGACCCGTG GACAGTATGG TGACATTCGG TAAACGGTA TTCNTTACTA  
 459 P C N S P Q P L R Y N Q A L M Q I V Q A G T M G S Y P M K G Y T I I

2101 TTCATNAAGA AGGATTCCAA ATGACCATAC ATGGACTTA ATGCTGAATT TATTACTACG AATGGCTTCG GGGCAGTCC ACTTCACCGG CAGCTTTATT  
 AAGTATTTCT TCCTAAGGTT TACTGGTATG TAGCCTGAAT TACGACTTAA ATAAATGATG TTACCGAAGC CCGCGTCAGG TGAAGTGCC GTCGAATAA  
 425 E Y L L I G F S W V D S K I S F K N S R I A E P A T W K V P L K I

2201 TCGTGTCTAG ATTCTATAGT GTCTTCATTA TCTACCTTAA AACTCTGGC AAGTCCAAA TCTGCTACTT TGTAGATATT ATGTTACCA ACAGGAGAT  
 AGCACAGATC TAAGTATCTA CAGAAATAT AGATGGAATT TTTGAGACCG TTCAGGTTTT AGACGATGAA ACATCTATAA TACAAGTGGT TGCTCCTGTA  
 392 E H R S E Y I D E N D V K F V R A L G F D A V K Y I N H E G V L V

2301 TTCTGGCAGC CAGATCTCTG TGAATGTAGT TCCGAGACTC CAGATAGGCC ATTCCAGAGG CAACCTGTGC GCCATGTCT ACCTGTTGAG TCAGATGGAT  
 AAGACCGTGC GTCTAGAGAC ACTTACATCA AGGCTCTGAG GTCTATCCGG TAAAGTCTCC GTTGGACACG GCGGTACAGA TGGACAACATC AGTCTACCTA  
 359 N R A A L D R H I Y N R S E L Y A M G S A V Q A A M D V Q Q T L H I

FIGURE 5D

2401 TTTTGAATCCA GTGTCAATTTT GGAGATATATTC TTGCAGACTT CCATGTCTCA TCAACTCTGT AATAATATAA ATTGGATCTT CTAAGTGCA AACAGATAA  
 325 K S G T D N Q L Y E Q L S G H R M L E T I I Y I P D E L T C V A Y  
  
 2501 AGCTGGATAA GCTTTGGATG TCTTAGGTTT TTTATATATCT GTCCCTCCCT CAGGAAGTCA TTTGGATCCA TTGAACCTGG TTTTAATGTT TTCACTGCTA  
 292 L Q I L K P H R L N K M I Q A E R L F D N P D M S G P K L T K V A  
  
 2601 CTGGAGTGGT ATGTGTTCCAC AGACCTTCCC ATACTTGGCC AACTGACCA GATCCCAATC GCTTCAGAAG CTGTATGGAG TTGCGGTCTA TCTCCCATTTG  
 259 V P T T N N W L G E W V E G F Q G S G L R K L L Q I S N R D I E W Q  
  
 2701 GTCCACGGTT TTATACGACA ATCAAAATGG AGCTGGGACC TGGATCTTTA AGCATGGTTT CCCAGCTTG ACACACAGGC CGTCACTTGT CTTCGGTGTAG  
 225 D V T K Y S L D F P A P V Q I K L C P K G L K V C L G D S T K T Y  
  
 2801 TGGCTCACAA ATTGTTTCAG TGTGAAAG ATTCTCTTC GGTGAGAA AATCCCTCT TCATCCAGTC TTTTAATTT GTAGTGTTT ACAACTGCTC  
 192 H S V F E N L T S F I R R R T L F F G G E D L R K I R Y H K V V A  
  
 2901 CATCTAAAAC TGAAGAGAG AATCTCTCTT TTTGGCTTTC ACTTCTCTG ATTAGAAAG AACCGTCTT GTTTTCTGAA TATAATAGTT GTTCTCTGTC  
 159 G D L V S L S F E G K Q S E S E R I L F S G T K N E S Y L L Q K E A  
  
 3001 ATCTGATCTT CCGATTGCTC CAAAGAACCA CGGCTCTGCC TGTAGGCTTC TGTCTCAGC CACGTAGTTA GAAGGAATAT AGCCTTGTAG TTGCTGACTG  
 125 D S R G I A G F F W P E A Q L S R D E A V Y N S P I Y G Q L Q Q S  
  
 3101 GAGCCATCTC GTCTTTTCTC CAAGTGCTG GCAMCCACC AGCCCTCATG CAAAGTGTC AGAAGTCAA GTTTGTACC TGCTCGGAAG CTCAAAGTCTC  
 92 S G D R R K E L H R A F W W G E H L T D L V Q L K D G A R F S L D

FIGURE 5E

3201 CAGCAGTCCG AGCCTGGTAA TCNACAAAG CCACAAAGTA GTGGCCATGC CTCCTGACT GGGGAGGCA AAGGCCCTT GGATTTTCAA TCACGGTTGA  
 GTCGTCAGGC TCGGACCAAT AGTTTGTTC GGTGTTTTC CACCGGTACG GAGACACTGA CCCTCTCGT TTCCCGGGA CCTAAAGTT AGTCCCACT  
 59 E A T R A Q Y D F L A V F Y H G H R Q S Q P S C L A G P N E I V T S

3301 CTTGTCTGCC TCCGTGGACA AACAGGGAG ATAGGGTTCT AGGTACTCCC AGAGCTCTG ACAGATGTTG CTCATTGTCC CTTGCTGGG AGAAGAGGAG  
 GNACAGACGG AGGCACCTGT TTGTCCCTC TATCCCNAGA TCCATGAGGG TCTCGAGAC TGCTACAAAC GAGTAACAG GNACCACCCC TCTTCTCTC  
 25 K D A E T S L C P L Y P E L Y E W L R Q C I N S M

3401 CAGGGCTTCT CCTCTCCCC TTAGTCTCTG CGATCCACCT TATCTTCTTT CACCAGCAA CTTTGAAGTC AGCACCAACT CACCATACTT CGGAGAGTAT  
 GTCCCGNAGA GGGAGAGGG AATCAGAGAC GCTAGGTGGA ATAGAAGGNA GTGTCTCTT GAACCTTCAG TCGTGGTTGA GTGGTATGAA GCCTCTCATA

3501 GCAAAGTCCC GTTTCAGATC AGTCCAGCAG CTGGGTGCA GCAAGTCTTA CTTACCGGCT TGCTTCTGT GGCTGGAGGT GCTACCCCGA  
 CGTTTCAGGG CAAAGTCTAG TCAGGTCTG GACCCAACGT CGTTCAGGAT GGACCTCTCT GAAATGCGCA AGGAAAGACA CCGACCTCCA CGATGGGCT

3601 GGCRAAAGT AGCAGGAGCT GGCAGCTGC TCACTAGGAA GGTGTCTTTT CTCTTATCT GCTTAAAGAT CCCACAACAA AATAAATA AATTTAAAG  
 CCGTTTGTGAC TCGTCTCGA CCCGTGAGC ACTGATCTT CCACAGAAA GNAGAATAGA CGAATCTTA GGTGTGTGT TTTATTTTAT TTTAATTTTC

3701 GGCTTTATTT AGACAAATAT CTGAGNACAG AATGGTGCCA TCTTGCCCTT TGTCCTAATA AAGATTAGC AAGAGGAGC TACTAACCCC TGGTAAACCC  
 CCGAATATA TCTGTTTATA GACTCTGTG TTAACACGGT AGNACGANA ACAGGTTAT TTTCAATCG TTCTCTTCCG ATGATTGGG ACCATTTTGG

3801 TCCACGTCTT GCTTTCGCCA GGTGCACTC GAGGATCTT CCATACCTAC CAGTCTGCG CCGCCGCGA CTCTAGAGTC GACCTGCAGA  
 AGGTGCAGAA CGAAGCGGT CCCAGCTGAG CTCCCTAGAA GGTATGGATG GTCAAGACGC GGACCTCCAG CCGCCGCGCT GAGATCTCAG CTGGACGTCT

3901 AGCTTGGCCG CCAATGGCCCA ACTTGTATT TGCAGCTTAT AATGTTTACA AATAAGCAA AATTTCACA ATAAAGCAT TTTTTCACCTG  
 TCGNACCGC GGTACCGGT TGNACNAATA AGTCTGAATA TTACCAATGT TTATTTCTGT TTAAGTGT TATTTCGTAA AAAAGTGAC

19/35

FIGURE 5F

4001 CATTCCTAGTT GTGGTTTGTC CAAACTCATC AATGTATCTT ATCATGTCTG GATCGGGAAAT TAATTCCGGCG CAGCACCATG GCCTGAAATA ACCTCTGAAA  
 GTAAGATCAA CACCAACAG GTTTGAGTAG TTACATAGNA TAGTACAGAC CTAGCCCTTA ATTAAGCCGC GTCTGGGTAC CGGACTTTAT TGGAGACTTT

4101 GAGGAACCTG GTTAGGTACC TTCTGAGGCG GAAAGAACCA GCTGTGGAAAT GTGTGTCACT TAGGGTGTGG AAGTCCCA GACTCCCCAG CAGGCAGAAG  
 CTCCTTGAC CAATCCATGG AAGACTCCGC CTTTCTTGGT CGACACCTTA CACACAGTCA ATCCACACC TTTCAGGGGT CCGAGGGGTC GTCCGTCTTC

4201 TATGCAAGC ATGCATCTCA ATTAGTCAGC AACCAGGTGT GGAAGTCCC CAGGCTCCCC AGCAGGCAGA AGTATGCMAA GCATGCATCT CAATTACTCA  
 ATACGTTTCG TAGCTAGAGT TAATCAGTGG TTGGTCCACA CCTTTCAGGG GTCCGAGGGG TCGTCCGTCT TCATACGTTT CGTACGTAGA GTTAATCAGT

4301 GCAACCATAG TCCCGCCCT AACTCCGCCC ATCCCGCCCC TAACTCCGCC CAGTTCGCC CATTCGCCG CCCATGGCTG ACTAATTTTT TTTATTTATG  
 CGTTGGTATC AGGCGGGGA TTGAGGCGG TAGGCGGGG ATTGAGGCGG GTCAAGGCGG GTAAGAGCGG GGTACCGAC TGATTAAAAA AATAAATATC

4401 CAGAGCCGA GGCGCCCTCG GCCTCTGAGC TATTCAGAA GTAGTGAGGA GGCTTTTTTG GAGGCCTAGG CTTTTCGAAA AAGCTGTAA CAGCTTGGCA  
 GTCTCCGGCT CCGCGGAGC CGGAGACTCG ATAAGTCTT CATCACTCCT CCGAAAAAAC CTCCGATCC GAAACGTTT TTCGACNAIT GTCGAACCGT

4501 CTGGCCGTCG TTTTACAAG TCGTCACTGG GAAACCCCTG GCGTTACCCA ACTTAATCGC CTTCAGCAGC ATCCCCCTTT CGCCAGCTGG CGTAATAGCG  
 GACCGGCAGC AATAATGTTG AGCACTGACC CTTTGGGAC CGCAATGGGT TGAATTAGCG GAACGTCTG TAGGGGAAA GCGGTCCGACC GCATTTATCGC

4601 AAGAGGCCCG CACCGATCGC CCTTCCCAAC AGTTGCGCAG CCTGATGGC GATGCGGCC TGATGGGTA TTTTCTCCTT ACGCATCTGT GCGGTATTTT  
 TTCTCCGGC GTGGCTAGCG GGAAGGGTTG TCAACGCGTC GGACTTACC GCTTACCGCG ACTACGCCAT AAGAGAGGAA TGGGTAGACA CGCCATNAG

4701 ACACCGCATA CGTCAAAGCA ACCATAGTAC GCGCCCTGTA GCGGCGCATT AAGCGCGCG GGTGTGTGTGG TTACGCGCAG CGTGACCGCT AACTTGGCA  
 TGTGGCGTAT GCAGTTTCTGT TGGTATCATG CGCGGACAT CGCCGCGTAA TTCCGCGCCG CCACACCAACC AATCGCGTC GCACTGGCGA TGTGAACGGT



20/35

FIGURE 5G

4801 GCGCCCTAGC GCGCGCTCCT TTGCGTTTCT TCCCTTCCTT TCTCGCCACG TTGCGCGGCT TTCCCGCTCA AGCTCTAAAT CGGGGCTCC CTTTAGGGTT  
CGCGGGATCG CGGGCGAGGA AAGCGAAGA AGGCMGGA AGNGCGTGC AAGCGGCCGA AAGGGCAGT TCGAGATTTA GCGCCGAGG GAATCCCA

4901 CCGATTAGT GCTTTACGGC ACCTCGACCC CAAAAACTT GATTGGGTG ATGTTTACG TAGTGGCCA TCGCCCTGAT AGACGGTTT TCGCCCTTTG  
GGCTNAATCA CGAAATGCG TGGAGCTGG GTTTTGTAA CTAAACCCAC TACCAAGTGC ATCACCCTT AGCGGACTA TCTGCCAAA AGCGGNAAC

5001 ACGTTGAGT CCAGTTCTT TAATAGTGA CTCTGTTC AACTGGAAC AACACTCAAC CCTATCTGG GCTATCTTT TGATTTATAA GGGATTTTGC  
TGCAACCTCA GGTGCAAGAA ATTATCACCT GAGAACNAGG TTTGACCTTG TTGTGAGTTG GGATAGAGCC CGATAAGAAA ACTAAATATT CCTTAAACG

5101 CGATTTCGGC CTATTGGTTA AAAATGAGC TGATTTAACA AAATTTAAC GCGAATTTTA ACAATATTT AACGTTTACA ATTTTATGTT GCATCTCAG  
GCTAAAGCCG GATAACCAAT TTTTACTCG ACTAAATTTG TTTTAAATG CGCTTAAAT TGTTTATAA TTGCAATGT TAAATATCCA CGTGAGAGTC

5201 TACAATCTGC TGTGATGCC CATAGTTAAG CCAGCCCGA CACCCGCTGA CGCGCCCTGA CCGCTTGTG TGTCCCGG ACAGAGGCGG TAGGCGAATG  
ATGTTAGACG AGACTACGG GTATCAATTC GGTGCGGCT GTGGCGGTT GTGGCGACT GCGCGGACT GCGCGAACAG ACAGAGGCGG TAGGCGAATG

5301 AGACAAGCTG TGACCGTCTC CGGAGCTGC ATGTGTCAGA GGTTTTCACC GTATCACCG AAACGCGGA GACGAAGGG CCTCGTGATA CGCCTATTTT  
TCTGTTGAC ACTGGCAGG GCGCTCGACG TACACAGTCT CCAAAAGTGG CAGTAGTGGC TTGCGCGCT CTGCTTTCCC GGAGCACTAT GCGGATANA

5401 TATAGGTTAA TGTATGATA ATATGGTTT CTAGAGCTC AGGTGGCCT TTTGCGGGA ATGTGCGCG AACCCCTATT TGTATTATT TCTAATATA  
ATATCCAAAT ACAGTACTAT TATTAGCANA GAATCTGCAG TCCACCTGA AAAGCCCTT TACACGCGC TTGGGATANA ACNAATANA AGATTATGT

5501 TTCAAATATG TATCCGCTCA TGAGACAATA ACCCTGATA ATGCTTCAAT AATATTGAA AAGGAAGAT ATGAGTATTC AACATTTCG TGTGCGCCTT  
ANGTTTATAC ATAGGCGAGT ACTCTGTTAT TGGNCTATT TACGAAGTAA TTATAACTTT TTCTTCTCA TACTCATAG TTGTAAAGG ACAGCGGNA

21/35

FIGURE 5H

5601 ATTCCCTTTT TTGCGGCATT TTGCCCTTCCT GTTTTGTGTC ACCCAGAAAC GCTGGTGAAA GTAAAGATG CTGAAGATCA GTTGGGTGCA CGAGTGGGTT  
 TAAAGGAAAA AACGCCGTAA AACGCCGTAA AACGGAAGGA CAAAACGAG TGGGTCTTTG CGACCACCTTT CATTCTCTAGT CAACCCACGT GCTCACCCAA

5701 ACATCGAACT GGAATCTCAG AGCGGTAAAG TCCTTGAGAG TTTTTCGCCC GAAGACGTT TTCCAATGAT GAGCACTTTT AAAGTTCTGC TATGTGGCGC  
 TGTAGCTTGA CCTAGAGTTG TCGCCATTCT AGGAACCTCTC AAAAGCGGG CTCTCTGCAA AAGTTACTA CTCGTGAAA TTCAAGACG ATACACCCCG

5801 GGTATTATCC CGTATTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAGTCAC AGAAAGCAT  
 CCATAATAGG GCATAACTGC GCGCGTTCT CGTTGAGCCA GCGCGTATG TGATAAGAGT CTTACTGAAC CAATCATGA GTGGTCAGTG TCTTTTGGTA

5901 CTTACGGATG GCATGACAGT AAGAGAAATTA TGCAGTGTG TGCAATACCAT GAGTGATAAC ACTGCGCCA ACTTACTTCT GACAACTATC GGAGGACCGA  
 GAATGCTTAC CGTACTGTCA TTCTCTTAAT ACGTCACGAC GGTATTGGTA CTCACTATTG TGACGCGGT TGAATGAAGA CTGTTGCTAG CCTCTGGCT

6001 AGGAGCTAAC CGCTTTTGTG CACAACATGG GGGATCATGT AACTCGCCTT GATCGTTGG AACCGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA  
 TCCTCGATTG GCGAAANAAC GTGTGTATCC CCTAGTACA TTGAGCGGAA CTAGCAACCC TTGGCCTGGA CTTACTTGG TATGGTTTC TCTCGCACT

6101 CACCACGATG CCTGTAGCAA TGGCAACAAC GTTGGCGAAA CTATTAACTG GGAACACT TACTCTAGCT TCCCGCAAC AATTAAATAGA CTGGATGGAG  
 GTGGTGTCTAC GGACATCGTT ACCGTGTGTTG CACCGCGTTT GATAATTGAC CGCTTGATGA ATGAGATCGA AGGCGCGTTG TTAATTATCT GACCTACCTC

6201 GCGGATAAAG TTGCAGGACC ACTTCTGCGC TCGGCCCTTC CGGCTGGCTG GTTTATTGCT GATAAATCTG GAGCCGCTGA GCGTGGTCT CCGGTATCA  
 CGCCTATTTC AACGTCTTGG TGAAGACGCG AGCCGGAG GCGGACCGAC CAATAACGA CTATTTAGAC CTCGGCCACT CGCACCCAGA GCGCCATAGT

6301 TTGCAGCACT GGGGCCAGAT GGTAAAGCCCT CCCGTATCTT AGTTATCTAC ACCAGCGGGA GTCAAGGCAC TATGGATGAA CGAATAGAC AGATCGCTGA  
 AACGTCTGTA CCGCGGTCTA CCATTCGGGA GGGCATAGCA TCAATAGATG TGCTGCCCCCT CAGTCCGTTG ATACCTACTT GCTTTATCTG TCTAGCGACT

22/35

FIGURE 51

6401 GATAGGTGCC TCACTGATTA AGCATTGGTA ACTGTCAGAC CAAGTTTACT CATATATACT TTAGATTGAT TTAAACTTC ATTTTAAAT TAAAGGATC  
CTATCCACGG AGTGACTAAT TCGTAACCAT TGACAGTCTG GTTCAAAATGA GTATATATGA NATCTAACTA AATTTGAAG TAAAAATTAA ATTTTCCTAG

6501 TAGGTGAAGA TCCCTTTTGA TAATCTCATG ACCAAATCC CTTAACGTGA GTTTTCGTTT CACTAGCGT CAGACCCCGT AGAAAGATC AAAGGATCTT  
ATCCACTTCT AGGAAAAGT ATTAGAGTAC TGGTTTTAGG GAATGCACT CAAAAGCAAG GTGACTCGCA GTCTGGGGCA TCTTTTCTAG TTCTCTAGAA

6601 CTTGAGATCC TTTTCTTCTG CCGTAATCT GCTGCTTGCA AACAAAAA CCACCGTAC CAGCGTGGT TTGTTTGGC GATCAAGAGC TACCAACTCT  
GAACTCTAGG AAAAAAGAC GCGCATTTAGA CGACGAACGT TTGTTTTTTT GGTGGCGATG GTGCCACCA AACAAACGGC CTAGTTCTCG ATGTTTGAGA

6701 TTTTCCGAAG GTAATGGCT TCAGCAGAGC GCAGATACCA AATACTGTTT TTCTAGTCTA GCGTAGTTA GGCCACCACT TCAAGAACTC TGTAGCACCG  
AAAAGGCTTC CATGACCGA AGTCGCTCTG CGTCTATGGT TTATGACAAG AAGATCACAT CCGCATCAAT CCGTGGTGA AGTTCTTGAG ACATCGTGCG

6801 CCTACATACC TGGCTCTGCT AATCTGTGTA CCAGTGGCTG CTGCCAGTGG CGATAAGTGG TGTCTTACCG GGTGGACTC AAGACGATAG TTACCGGATA  
GGATGTATAG AGCGAGACGA TTAGGACAAT GGTCAACCGAC GACGGTCACC GCTATTTCAGC ACAGATGGC CCAACCTGAG TTCTGCTATC AATGGGCTAT

6901 AGCGCCAGCG GTCCGGCTGA ACGGGGGTT CGTGCACACA GCCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGTAC CTACAGCGTG AGCTATGAGA  
TCCCGGTCCG CAGCCCGACT TGCCCCCAA GCACGTGTGT CCGGTGGAAC CTCGCTTGGT GGAJGTGGCT TGACTCTATG GATGTGGCAG TCGATACTCT

7001 AAGGCCACG CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGCTCGG AACAGGAGAG CGCAGGAGGG AGCTTCCAGG GGGAAACGCC  
TTCCGGGTGC GAAGGGCTTC CTTCTTTTCCG CTTGTCCATA GGCATTTCCG CGTCCACGC TTGTCTCTC GGTGCTCC TCGAAGGTCC CCTTTTGGCG

7101 TGGTATCTTT ATAGTCTCTGT CGGTTTTCGC CACTCTGAC TTGAGCGTGG TTGCTCTCTC GCAGGAGAG CCGGCGGAG CCTATGAAA AACGCCAGCA  
ACCATAGAAA TATCAGGACA GCCCAAGCG GTGAGAGCTG AACTCGCAGC TAAAAACT ACGAGCAGTC CCCCCGCTC GGATACCTTT TTGCGGTCTGT

FIGURE 5J

7201 ACGCGCCCTT TTACGGTTC CTGGCCTTTT GCTGGCCTTT TGCTCACATG TTCTTTCCCTG CGTTATCCCC TGATTCTGTG GATAACCGTA TTACCGCCTT  
TGCGCCGGAA AAATGCCNAG GACCGGAAJAA CGACCCGMAA ACGAGTGTAC AAGAAAGGAC GCAATAGGGG ACTATGGCAT AATGGCGGAA

7301 TGAGTGAGCT GATACCGCTC GCGCAGCCG AACGACCGAG CGCAGCGAGT CAGTCAGCGA GGAAGCGGAA GAGCGCCAA TACGCAAAAC GCCTCTCCCC  
ACTCACTCGA CTATGGCGAG CGCGCTCGGC TTGCTGGCTC GCGTGGCTCA GTCACTCGCT CCTTGGCCTT CTGCGGGTT ATGCGTTTGG CGGAGAGGGG

7401 GCGCGTTGGC CGATTCAATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG GAAAGCGGC AGTGAGCGCA ACGCAATTAA TGTGAGTTAG CTCACTCATT  
CGCGCAACCG GCTAAGTAAT TACGTCGACC GTGCTGTCCA AAGGCTGAC CTTTGGCCCG TCACTCGGT TCGTTAATT ACACTCAATC GAGTGAGTAA

7501 AGGCACCCCA GGCCTTACAC TTTATGCTTC CGGCTCGTAT GTTGTTGGA ATTGTAGCG GATAACAATT TCACACAGGA AACAGCTATG ACATGATTAC  
TCCGTGGGGT CCGAATGTG AATACGAAG GCCGAGCATA CAACACACCT TAACACTCGC CTATTGTTAA AGTGTGCTT TTGTCGATAC TGTACTAATG

7601 GAATTAA  
CTTAATT

24/35

1 GCGGCCGCAG AGAAGCAGA GGATGGGGCT TAGCAGCTGG CAGAGCCAGG ACCGGGGAGG TAGCAGAAAG ACCACAAGTA CAAGAAGTC CTGAACCTT  
CGCCGGGGTC TCTTTCGTCT CCTACCCCGA ATCGTCGACC GTCGCGGTCC TCGCCCTCC ATCGTCTTTC TCGTTCAT GTTCTTCAG GACTTTCANA

101 GGTTTTCTG CTGCAGCCCA TTGAGAGTGA CGACATGGAG CACAGACCC TGAAGATCAC CGACTTTGCC CTGSCCCGAG ACTGGCACAA AACCCACAA  
CCAAACGAC GACGTCGGGT AACTCTACT GCTGTACCTC GTGTTCTGGG ACTTCTAGTS GCTGAACCG GACCGGGCTC TCACCGTGT TTGGTGTGT

201 ATGAGTGCCG CXGGCACCTA CXCTGGATG GCTCCTGAGG TTATCAAGG CTCCACCTTC TCTAAGGCA GTGAGCTCTG GAGTTTGGG GTGCTGCTGT  
TACTCACGGC G?CCGTGGAT G?GGACCTAC CGAGGACTCC AATAGTTCCG GAGGTGAAG AGATTCCCT CACTGCAGAC CTCAAAACCC CACGACGACA

301 GCGAACTGCT GACCGGGGAG XTGCCATACC GTGGCATTGA CTGCCCTGCT GTGGCCTATG GCGTAGCTGT TAACAGCTC AACTGCCAT CCATCCACCT  
CCCTGACGA CTGGCCCTC ?ACGGTATGG CACCGTAAC TACCGTAAC CACCGATAC CCGATCGACA ATTGTTGAG TGTACGGTA GTTAGGTGA

401 GGCC  
CCGG

FIGURE 6

25/35

FIGURE 7A

1 ATGAGAGCGT TGGGCGGCA CCGCGCGAG CTGCGCTGCG TGGTGTGTTTT TTCTGCAAGG ATATTGCGA CTATTACAAA TCAAGATGCTG CTGTGTGATCA  
TACTCTGCA ACCGCGCGCT GCGCGCGTGC GACGCGCGAG AGCAATAAAA AAGAGCTTAG TATAAAGCGT GATAATGTTTT ACTCTAGAG CGACACTAGT

101 AGTGTGTTTT AATCAATGAT AAGAACAAATG ATTGATCAGT GGGAACTGA TCATCATATC CCATGCTATC AGATCCGCG GAGAGCGTGG GHTGTGCHTT  
TCACACAAAA TTACTTAGTA TTCTTTGTTAG TAAGTACTCA GCGCTTCACT AATACTATAG GGTACCATAG TCTTAGGCGC CTCTGCGAGC GCACCGGCAA

201 GAGACCCCG AGCTCAGGCA CAGTGTACGA AGCTGTGCT GTGCAAGTGC ATGTATCTGC TTGATCACA CTGCAAGTGC TGGTGGATGC CCCAGGGGAG  
CTCTGCGGTC TCGAGTCCCT GTACATGCT TCGACGCGCA CAGCTTCAGC TACATAGAG AGGTAGTGT GAGCTTACG ACCAGCTAGC GGGTCCCTTG

301 ATTCTGTC TCTGCTCTT TAAACACAG TCGCTGAATT GGTAGCAGA TTTGATTTA CAAGACAG GAGTGTCTTC CATGCTCAT TTCAAAATGA  
TAAAGACAG AGACGAGAA ATTGCTGCG AGGAGCTTAA GGTGCTGCT AAGCTAAT AAATCTATC CTCAACAAAG GTACCGATAA AACTTTTACT

401 CAGAAACGCA AGCTGAGAA TACCTACTTT TTATTACAG TCAAGCTATC AATTAACAAA TATTGTTTAC AGTGAATATA AGAAATAGC TGTGTTTACG  
GTCTTTGGT TCGACCTTT ATGATGAAA AATAAGTCTC ACTTGATG TTAAATGCTT ATAACAATG TCACTCATAT TCTTTATGCG ACGAATGTC

501 ATTAAAGAGA CTTTACTTTA GAAAAATGGA AAGCCAGGAG GCGCTGCTCT GCATATCTCA GAGCGTTCCA GAGCGGATCC TGGATGCGT GCTTTGCGAT  
TAATCTTCT GAAATGAAT CTTTTTACCT TTTGCTCTG CGGACGAGA GGTATAGACT CTGCGAAGCT CTCGCTTAGG ACCTTACCCA CGAAACGCTA

601 TCACAGGGGG AAGCTGTAA AGAAGAAAT GAGCTGTTG TTAAAAAGGA GGAAGAGTG CTTGATGAAT TATTTGGAG GCACATAAG TGTGTGTGCA  
AGTGTGCGG TTTGACAT TCTTCTTTCA GGTGACAGC AATTTTCT CTTTTTCA GAGTACTTA ATAAAGCTG CCTGTATTCC AGGACAGGT

701 GAAATGAAT GGGAGGGA TGCACAGG TGTTCACAAAT AGATCTAAT CAACTGCTC AGACCACTT GGCACATTA TTTCTTAAAG TAGGGAAC  
CTTAATTTCA GCGTGGCTT AGTGTGCGC AAGATCTTA TCTAGATTTA GTTGAAGAG TCTGTGTAA GGTGTTAAT AAGAAATTC ATCGCTTTC

FIGURE 7B

801 CTTATGAGATA AGGTGCAAAAG GTGTTTCATCT GAACCATGGA TTGCGGCTCA CCTGGAATT AGAAGACAAA GCACCTCGAG AGGCGAAGCTA CTTTGAGATG  
 GAATACCTAT TCCAGGTTTC GACAAGTACA CTTGCTATCT AAGCCCGAGT GACCCCTTAA TCTTTTGTTC GTGAGCTCC TCCCGTTGAT GAAACTCTAC

901 ACTACCTTATT CAACAAACAG AACTATGATA CGGATCTCTT TTGCTTTTCT ATCATCTACT GCAAGAAACG ACACCGATA CTACACTTGT TCCCTTTCAA  
 TCAATGATAA GTTGTGTGTC TTGATACTAT GCTTAAGACA AACGAAACA TACTAGTCAC CTTTCTTTTC TGTGCTAT GATGTGAAC AGGAGAAAGTT

1001 AGCATCCGAG TCAATCAAGT TTGGTTAGCA TCGTAGAANA GGGATTATTA AATGCTACCA ATTCAAGTCA AGATTATGAA ATTGACCAAT ATGAAAGAGTT  
 TCGTAGGCTC ACTTAGTCCA AACCAATGGT AGCATCTTTT CCTAAGATAT TTACGATGCT TAAGTTCACT TCTAATACTT TAAGTGTTA TACTTCTCAA

1101 TTCTTTTTCT GTCAGGTTTA AAGCCTACCC ACAATACAGA TGTACTGCA CTTTCTCTCG AATATCATTT CTTTCTGAGC AAGGGGTCT TCAATACCGA  
 AACAAAAGCA CACTCCAAAT TTGCGATGGG TGTTAGTCT ACATGACCT GGAAGAGAGC TTTTAGTAAA GGAACACTCG TTTTCCGACA ACTATTGCT

1201 TACAGCATAT CCAAGTTTTC CAATCATAG CACCAACAG CAGATATAT ATTCCATGCA GAATATGAG ATGCCCAATT TACCAAAATG TTCACGGTCT  
 AATGCGTATA GTTCAAAAC GTTAGTATTC GTGCTGCTG CTCTTATATA TAAGGTACTT CTTTACTAC TACGGTTTAA ATGGTTTTAC AAGTGGGACA

1301 ATATAGAAAG GAACCTCMA GTGCTCGCAG AAGCTTGCGC AAGTCAAGCG TCCGTCTTCT CGGATGGATA CCGATTACCA TCTTGGAGCT GGAAGAAAGTG  
 TATATTCTTC CTTTGAATT CAGGAGCTC TTGGAAGCG TTCACTGCGC AGGACAAAGA GCTACCTAT GGTAAATGCT AGAACCTGGA CTTTCTTCAC

1401 TTCAACAAAG TCTGCCAAT GCACAGAGA GATCAGAAA GAACTCTGGA ATAGAAAGC TAACAGAAA GTGTTTTGAC AGTGGTCTC GAGCACTACT  
 AAGTCTGTC AGAGGTTCA CTTCTCTCT CTAGTCTCTT CTTAGAGCT TATCTTTTCC ATTGCTTTT CACAAAGCTG TCACCCACAG CTCGTCTATGA

1501 CTAAACATCA GTGAAGCCAT AAGAGGTTTC GTGCTCAAGT GGTGTCATA CAATTCCTCT GGCACATCTT GTGACAGGAT CTTTTAAAC TCTCCAGGCC  
 GATTTCTACT CACTTCGGA TTTTCCCAAG GATCACTTCA GCACAGCTAT GTTAAGGAA CCGTAGAA CACTCTGTA GGAATTTG AGAGGTCGG

FIGURE 7C

1601 CTTTCCTTT CATCCAGAC AACATCTAT TCTATGCAAG AATTGTTT TGTCTCTCT TCTATGTCCT TTTAACCTG CTAATTTGTC ACNAGTACAA  
 CGAAGGHAHA GTAGGTTCTG TTATAGATA ACATAGTTTC TTAACACAA ACAGAGHHA AATTGGHAC GATTAAGACG TGTTCATGTT

1701 AAGCAATTT AGGTATGAA GCCAGCTACA GATGCTACAG GTGACGGGAT CCTCAGATAA TGAGTACTTC TACGTTGATT TCAGAGATA TGAATATGAT  
 TTTCTTAA TGCATACTTT GGTGCTATCT CTACATATCT GATGCTATCT GAGTCTATT ACTCATGAG ATGCAACTAA ACTCTCTTAT ACTTATACTA

1801 CTCAAATGCG AGTATCCAG AGAAATTTTACAGTTTGGG AGGTACTAGG ATCAGGTGCT TTTGGAAGG TGATGAAGG AACAGCTTAT GGAATTAGCA  
 GAGTTTACCC TCNAGGTTG TCTTTTAAAT CTCNAAACCT TCCATGATCC TAGTCCACGA AACCTTTTC ACTACTTGGG TTCTCGAATA CCTTAATCGT

1901 AACAGGAGGT CTCNATCCAG GTTACCCTCA AATGCTGAA AGAAGAGCA GACAGCTCTG AACAGAGGG ACTCATGTCA GAACTCAGA TCATGACCCG  
 TTTCTCTCA GAGTTAGGTC CAATGCCAGT TTTACGACTT TCTTTTCTG CTGTCGAGC TTTCTCTCG TGAGTACACT CTTGAGTTCT ACTACTGGGT

2001 GCTGGGAGG CACGAGATA TTGTGAGCTGTGGGGGGG TGCACACTGT CAGGACCAAT TTACTTGATT TTTGAATACT GTTCTATCG TGAATCTCTG  
 CGACCTTCTG GTCTCTTAT AACACTTGA CGACCCCGCG AGCTGTGCA GTCTGTGTTA ATGCACTAA AACCTTATCA CAAGATACC ACTAGAGAG

2101 AACATCTAA GAGTAAGAG AGAATTTT CACAGGACTT GCACAGACT TTTCAAGGA CACAAITTC GTTTTTACC CACTTCCAA TCACATCCAA  
 TTGATAGATT CTTCATTTTC TCTTTTAA GTTCTCTGAA CTGTCTCTTA AAGTTCTTT AAGTTCTTT CAAAATGGG GTGAAGGTT AGTGTAGGTT

2201 ATTCCAGCAT GCTGGTTCA AGAAGGTTG AGATACACCC GCACTGCGAT CAATCTCAG GCTTTCATGG GAATTCATTT CACTCTGAG ATGAATTTGA  
 TAACTCGTA CGGACCAAGT TCTCTTCAG TCTATGTGG CTTGAGCTC GTTAGTAA CTNAGTAA GTGAGACTTC TACTTTAACT

2301 ATATGAAAG CAAAAAGGC TGGAGAGAA GAGGAGCTTG AATGTCTTA CATTTGAGA TCTTCTTGG TTTGCATATC AGTTGCCAA AGGAATGGA  
 TATATTTTC GTTTTTCGG ACCTTCTCT CTCTCTGAC TTACACCAAT GTAACTTCT AGAAGAACG AACGTATAG TTCAACGGTT TCTTACCTT



FIGURE 7D

2401 TTTCTGGGAAAT TTAAGTCGTG TGTTCAACACA GACCTGGCCG CAGGGAACGT GCTTGTCCACC CACGGGAAGG TGGTGAAGAT ATGTGACCTT GGATTGGCTC  
 A AAGACCTTA ATTACAGCAC ACAAGTGCT CTGGACCGC GTTCCCTTCA CCAACAGTGG GTGCCCTTC ACCACTTCTA TACACTGAAA CCTAACCGAG

2501 GAGATATCAT GAGTGATTC ACHATGTTG TCAGGGGGCA TCCCGCTGTG CCGTAAAT CCGATGCCCC CGAAGCCCTG TTGAAAGCCA TCTACACCAT  
 CTCTATGTG CTCACTAAGG TTGATACAACT TTTTCCCTT ACGGGACAGC GCACATTTTA CCTACCGGGG GCTTCCGAC AACCTCCGT AGATGTGGTA

2601 TAAAGTGTAT GTCTGGTCAAT ATGCAATATT ACTGTGGCA ATCTTCTCAC TTGGTGTGAA TCCTTACCCCT GGCATTCCGG TTGATGCTAA CTTCTACAAA  
 TATTCTACTA CAGACCAGTA TACCTTATTA TCACACCCCT TACAAAGACTG AACCACTT AGGAATGGGA CCGTAAGGCC AACACTGAGT GAAGATGTTT

2701 CTGATTCAAA ATGATTTTAA ATGGATCAG CCATTTTATG CTACAGAGA AATATACATT ATAATGCCAAT CCGCTGGGC TTTCGACTCA AGAAGCGG  
 GACTAAGTTT TACCTAAATT TTACCTAGTC GGTAAAATAC CATCTCTTCT TTATATCTAA TATTACGTTA GGACGACCG ANACTGAGT TCCTTTGGCG

2801 CATCCTTCCC TAATTGACT TCGTTTTAG GATGTACCT GGCAGATGCA GAAGAAAGCA TGTATCAGAA TGTGATGGC CGTCTTCCG AATGTCCCTCA  
 GTAGCAAGG ATTAACTGA AGCAAAATC CTACAGTGA CCGTCTACGT CTCTCTGCT ACATAGTCTT ACACCTACC GACAAAGCC TTACAGGACT

2901 CACCTACCAA AACAGCGGAC CTTTCAGCAG ACAGATGGAT TTGGGGCTAC TCTCTCCGA GCGTCAGGTC GAAGATTCGT AGAGAACAA TTTAGTTTTA  
 GTGGATGGT TTGTCCGCTG GAAGTCGT TCTCTACCTA AACCCGATG AGAGAGGCGT CCGAGTCCAG CTTCTNAGCA TCTCTTGT AATCANAAT

3001 AGGACTTCAT CCGTCCACCT ATCCCTAACA GGCTGTAGT TACCAAAACA AGTTAATTT CATCACTAA AGAATCTA TTATCAGCTG CTGCTTCACC  
 TCCCTGAAGTA GGGAGGTGGA TAGGGATTCT CCGACATCTA ATGGTTTTGT TCAATTAAG CTAGTCATTT TCTTTTAGAT AATAGTTGAC GACGAAGTGG

3101 AGACTTTCT CTAGAGAGCG  
 TTCTGAAGAAT CATCTCTCC

FIGURE 8A

1 TCGGCTCCA CCGCCACAG GAGAGTCAGA CCTGKGGG GAGGGCCC CCAACTCAG TTGGATCCT ACCGAGTGA GCGCGGCCA TGGAGCTCC  
 AGCCGAGGT GCGCGGTCC CTCTCACTCT GACCCCCC GCTCCCGGG GATTGAGTC AAGCTAGGA TGGCTCACT CCGCGCGCT ACCTGAGGC  
 M E L R

101 GGTGCTCTC TGTGGGCTT GGTGCGCC AGCTTTGGA GAGACCTGC TGAACACAA ATTGAAACT GCTGATCTGA AGTGGTGAC ATTCCCTCAG  
 CCACGACGAG ACGACCGAA GCAACCGG TCAAACTTT TCTTGNGG ACTTGCTT TAACTTTGA CCACTAGACT TCACCCACTG TNAAGGAGTC  
 S V L L C W A S L A A A I E E T L L N T K L E T A D L K W V T F P Q

201 GTGACGGCC AGTGGGAGGA ACTGACGGC CTGATCAGG MACAGCACG GTGCGCAC TACGAGTCT GTGACGTGCA GGTGCCCC GGGCAGGGCC  
 CACTGCGG TCACCTCCT TGACTCGCG GACTACTCC TTGCTGTCT GACCGGTG ATGCTTACA CACTGCACT GCGACCGGC CCGTCCGG  
 J R V D G Q W E E L S G L D E E Q H S V R T Y E V C D V Q R A P G Q A H

301 ACTGGCTTG CACAGGTGG GTCCACGGC GGGGCGCT CCAGGTGAC GCGAGCTGC GCTTACCAT GCTGAGTGC CTGTCCCTGC CTGCGCTGG  
 TGACCGAAGC GTGTCCAAC CAGGTGCGT CCGCGGCA GTTCACTG CGGTCCGAG CGAGTGGTA CGAGTCAAG GACAGGACG GAGCCGACC  
 72 W L R T G W V P R R G A V H V Y A T L R F T H L E C L S L P R A G

401 GCGTCTCTC AAGGAGACT TCACGTCTT CTACTATGAG AGCGATGCG ACACGGCCAC GCGCTCACG CCAGCTTGA TGGAGAACCC CTACATCAAG  
 CCGAGAGCG TTCTCTTGA AGTGGCAGAA GATGATCTC TCGTACGCC TGTCCCGTG CCGGAGTGC GGTGAGCTT ACCTCTTGG GATGTAGTTC  
 105 R S C K E T F T V F Y Y E S D A D T A T A L T P A W H E N P Y I K

501 GTGGACACG TGGCGCGGA GCATCTCAC CGGAGCGCC CTGGGCGGA GGGCACCGG AAGTGAAATG TCAGACGCT GGTCTGGGA CCGCTCAGCA  
 CACTGTGCC ACCGCGGCT GTAGAGTGG GCTTCGCG GACCCGGCT CCGTGGGCC TTCCACTTAC AGTTCTGGA CCGAGACCT GCGGAGTCT  
 138 V D T V A A E H L T R K R P G A E A T Q K V N V K T L R L Q P L S K

601 AGGTGGCTT CTACCTGGC TTCCAGGACC AGGTGCTG CATGGCCCTG CTATCCCTGC ACTCTTCTA CAAGAATGC GCCCAGCTGA CTGTGACCT  
 TCCGACCGA GATGACCGG AAGTCTCTG TCCACGGAC GTACCGGAC GATAGGACG TGGAGAGAT GTTTTCAG CCGGTGAGT GACACTTGA  
 172 A G F Y L A F Q D Q G A C H A L L S L H I F Y K K C A Q L T V N L

30/35

FIGURE 8B

801 TACTGCGTG AGGATGGCCA GTGGGCGGAA CAGCGGTCA GGGCTGCA CTTGCTCCG:GGTTGAGG CAGTGAAG GAAACCAAG TGCCGAGCCT  
 ATGACGGCAC TCCTACCGGT CACCGGCTT GTTGGCACTT GCGGACGTG GACAGAGGC CCCAGCTCC GTGACTCC CTGTGGTTC AGCGTCGGA  
 218 Y C R E D G Q W A E Q P V T G C S C A P G F E A A E G N T K C R A C

901 GTGCCCCAGG CACCTTCAAG CCCCTGTCAG GAGAGGCTC GTGCCAGCA TGCCAGCCA ATAGCCACTC TAAACCAATT GATCAGCGG TCTGCCAGTG  
 CACGGGTCCC GTGGAAGTTC GGGGACATTC CTATTGTCAG GAGGTCGGT ACGGTTCGGT TATCGGTGAG ATTGTGTAA CCTAGTGGC AGACGTCAC  
 272 A Q G T F K P L S G E G S C Q P C P A N S H S N T I G S A V C Q C

1001 CCGCTCGG TACTTCCGG CAGGCACAGA CCCCCTGGT GCACCTGCA CCACCCCTCC TTGCGCTCG GAGGCTGG TTTCGGCCT GAAAGCTCC  
 GACGAGCCC ATGAAAGCCC GTGCGTGTCT GGGGCCCCA CTTGGGACT GTTGGGAGG AAGCCGAGG GCTGCGACC AAGGCGGA CTTCGCCAGG  
 305 R V G Y F R A R T D P R G A P C T T P P S A P R S V V S R L N G S

1101 TCCCTGCACC TGGATGGAG TGCCCCCCTG GAGTCTGGT GCCGAGAGA CCTCACCCTAC GCGCTCCCT GCGGAGTG CCGACCCGA GCGTCTGTG  
 AGGAGCTGG ACCTTACCTC ACGGGGAC CTCAGACCAC CGGCTCTCT GAGTGTGATG CCGGAGCGA CCGCCCTCAC GCGTGGCCT CCGAGGACAC  
 338 S L H L E W S A P L E S G G R E D L T Y A L R C R E C R P G G S C A

1201 CCGCCTCGG GGGAGACCTG ACTTTGACC CCGGCCCCG GAGCTGGTG GAGCCCTGG GAGCCTTGG TGCTGGTTC AGGCTACGT CCTGACTTCA CCTATACCTT  
 GCGGAGCCC CCGTCTGGAC TGAAACTGG GCGGGGGC CTTGAGCCAC CTCGGGAGCC ACCACCAAGC TCCCATGCA GACTGAAGT GGATATGAA  
 372 P C G G D L T P D P G P R D L V E P W V V V R G L R P D F T Y T F

1301 TGAGTCACT GCATTGACG GGTATCTTC CTTAGCCAG GGGCCGTC CATTGAGCC TGTCAATGTC ACCACTGACC GAGAGTACC TCCTGCGAGTG  
 ACTCCAGTGA GGTAACTTC CCGATAGAG GAATCGTGC CCGGGCAG GTAACTCG ACAGTTACAG TGCTGACTGG CTCTCATGG AGGAGTCA  
 405 E V T A L N G V S S L A T G P V P F E P V N V T T D R E V P P A V

1401 TCTGACATCC GGTGAGCG GTCTCACCC AGACCTTGA GCTGGCCTG GCGTGTCCC CCGGCACCA GTGGGCTGT CCGGACTAC GAGTCAANT  
 AGACTTAGG CCGACTGCG CAGGATGGG TCGTCAACT CCGACCGAC CCGCAJAGG GCGGCTGGT CACCCGACA CGACTGATG CTCCAGTTTA  
 438 S D I R V T R S S P S S L S L A W A V P R A P S G A V L D Y E V K Y

[illegible]

32/35

FIGURE 8D

2201 CCTCGGCTA AACGAGGAC AGTTCACAGT CATCCAGCTC GTGGCATGC TGGGGGCAT CCGCTCGGC ATCGGTACC TTCCGAGAT GAGCTACTC  
 705 L R L N D G Q F T V I Q L V G M L R G I A S G M R Y L A E M S Y V  
 2301 CACCGAGCC TGGTCTCTG CACATCTTA GTCMACCA ACCTGCTCTG CAAAGTGCT GACTTTGACC TTTCGGAAT CCTGGAGGAG AACTCTTCCG  
 GTGGCTCTG ACCGACGAGC GTTGTAGAT CAGTTTCGT TGGACAGAC GTTTCACAGA CTGAACCGG AAGGCTAA GAGCTCTC TTGAGNAGCC  
 738 H R D L A A R N I L V N S N L V C K V S D F G L S R F L E E N S S D  
 2401 ATCCACCTA CACGAGCTC CTGGAGGAA AGATTCCCAT CCGATGGACT GCGCGGAGG CCAATGCCCTT CCGAAGTTC ACTTCGCCA GTGATGCCG  
 TAGATTGGAT GTGCTGAGG GACCTCTCTT TCTAAGGTA GGCTACCTGA CGGGCCCTCC GGTACCGAA GCGCTCAAG TGAAGCCGT CACTACGAC  
 772 P T Y T S S L G G K I P I R W T A P E A I A F R K F T S A S D A W  
 2501 GAGTTACGG ATTGTGATG GGGAGGTGAT GTCTTTTGG GAGAGCCGT ACTGGACAT GAGCAATCAG GACGTATCA ATGCCATTGA ACAGGACTAC  
 CTCATGCCCC TAACTACTA CCTCCACTA CAGTAAACCC CTCTCGGCA TGACCTGTA CTGTTAGTC CTGCACTAGT TACGTAAT TGTCCTGATG  
 805 S Y G I V H W E V H S F G E R P Y W D H S N Q D V I N A I E Q D Y  
 2601 CGGCTGCCC CGCCCCAGA CTGTCCACC TCCTCCACC AGTCATGCT GACTGTGG CAGNAGACC GGAATGCCG GCGCCGCTTC CCCCAGGTG  
 GCGACGGG GCGGGCTCT GACAGGTGG AGGGAGTGG TCGAGTACGA CCTGACNACC GTCTTCTGG CCTACGGC CGGGCGNAG GGGTCCACC  
 838 R L P P P P D C P T S L H Q L M L D C W Q K D R N A R P R P P Q V V  
 2701 TCAGCGCCT GGACAGATG ATCCGACCC CCGCAGCCT CAATATCGT GCGCGGAGA ATGGCGGC CTCACACCT CTCTGGACC AGCGGAGCC  
 AGTCGGGGA CTGTTCTAC TAGGCTTGG GCGGTGGA GTTTAGAC CGGCCCCCT TACCGCCCC GAGTGTGGA GAGGACCTGG TCGCCGTGG  
 872 S A L D K M I R N P A S L K I V A R E N G G A S H P L L D Q R Q P  
 2801 TCACCTACTA GCTTTTGCT CTGTGGGGA GTGGCTTGG GGCATCAAAA TGGGAGATA CGAAGAAAT TTCCAGGCC CTGGCTTGG CTCTTCGAG  
 AGTCATGACT CGAAGCCGA GACACCCCT CACCGAGCC CGGTAGTTTT ACCCTTCTAT GCTTCTTCA AAGCTCGC GACCGNACC GAGGAGCTC  
 905 H Y S A F G S V G E W L R A I K M G R Y E E S F A A A G F G S F E

FIGURE 8E

2901 CTGTCAGCC AGATCTCTGC TGAGGACCTG CTCGTAATCG GATTCACCTCT GCGGACAC CAGAGAAA TCTTGCCAG TGTCCAGCAC ATGAACTCCC  
 GACCACTCGG TCTAGAGACG ACTCCTGAC GAGGCTTAG CTCAGTAGA CCGCCCTG GTCTCTTT AGAACGGTC ACAGTCTGT TACTTCAGGG  
 938 L V S Q I S A E D L L R I G V T L A G H Q K K I L A S V Q H M K S Q

3001 AGCCCAAGCC GGAACCCCG GTGGACAG GAGCACCGC CCGCAGTAC TGACCTGCG GAACTCCCA CCGCAGGAC ACCCCCTCC CATTTTCGG  
 TCCGTTCCG CCGTTGGGC CCACCTGTC CTCCTGGCG GCGCTCATG ACTGACGTC CTGAGGGT GCGTCCCTG TCGCGAGGG GTAAAGGCC  
 972 A K P G T P G G T G G P A P Q Y O P A G T P H P R D T A S P F S G

3101 GGCAGAGTGG GGACTCACAG AGGCCCCAG CCGTGTGCC CCGTGGATTG CACTTGAGC CCGTGGGTG AGGAGTTGCG AATTTGAGA GACAGATTT  
 CCCTCTACC CCGAGTGTG TCCGGGGTC GGGACACGG GGCACCTAAC GTGAACCTG GGCACCCAC TCTCAACG TTAACCTCT CTGTCTAAA  
 1005 A E W G L T E A P S P V P R W I A L O A R G V R S W Q F G E T G F

3201 GCGGTTCTG CCATMATAG AGGGANAT CACCCCCAG CCACCTGGR GAATCCAGA CCAAGGTGA GCGGCTTT CCGTCAGGAC TGGTGTGAC  
 CCCCCAGAC GGTATTATCC TCCCTTTTA GTGGGGTC GTGGAGCCC CTGAGGTCT GTTCCCACT CCGCGGANA GCGAGTCTG ACCACACTG  
 1038 G G S A I I G G E N H P P A T S G N S R P R V R A P F P Q D W V O P

3301 CAGAGGAAA GGAAGTCCC AACATCTCC AGCTCCCA GGTGCCCC TCACCTTGT GGTGGCTTC CCGCAGACA AGAGAGTGT GACTCCCTTG  
 GTCTCCTTT CTTACAGGG TTGTAGAGG TCGAGGGT CCACGGGG AGTCGACTA CCGACGANG GCGTCTGCT TCTCTACA CTGAGGAGC  
 1072 E E K E V P N I S Q P P Q V P P S P O W V R S R R P K R V O L P C

3401 CCAGTCCAG AGTGGGGGG CTGTCCACG GGCAGAGAG GGTGTGAG GCCAGTGAC AAATCATTT GCGTTGTAG TCCCACTTG CTGCTGTAC  
 GGTGAGTTC TCACCCCCC GACAGGGTCC CCGTTCTTC CCGACAGTCC CCGGTACTG TTTAGTATC CCGAACATC AGGTTGAGC GACGACAGT  
 1105 Q L Q S G G A V P G G K K G C Q G P V T K S L G F V V P T C C H

3501 CACCAACTC ATCATTTTT TTCCCTTCTA ATGCCCTC CCGAGTGC TCGCTTCTA TTGAGTTT TCGAGTTTG TTTTGTGT TAATTTTCT  
 GTGTTTGG TTAGTAAAA AAGGGACAT TTACGGAG GGGTGGAG ACCGAGTAT AACTTCCAA AACTCAAAAC AAAAACCAGA ATTAAAAAGA  
 1138 I I Q T Q S F F S L V N A P P P A A A F I L K V F E F C F W S O F F S

34/35

FIGURE 8F

3601 CCCCGTTCCC TTTTGTGTTTC TTGCTTTTGT TTTTCTACCG TCCTTGTGAT AACTTTGTGT TGGAGGGAAC CTGTTTCACT ATGGCTTCTT TTGCCCCAAGT  
GGGGCAAGG AANAACAAAG AAGCNAACA AANAATATGC AGGACAGTA TTGAACACAA ACCTCCCTTG GACAAAGTGA TACCGAGGA ACGGTTTCA  
1172 P F P F C F F V L F P Y R P C H N P V L E G T C F T H A S F A Q V

3701 TGAACAGGG GCCCATCATC ATGTCTGTTT CCAGAACAGT GCCTTGCTCA TCCACATCC CGGACCCCG CCTGGACCC CCNAGCTGTG TCCTATGAG  
ACTTTGTCCT CCGGTAGTAG TACAGACAAA GTCTTTGCTCA CGGACCAAGT AGGCTGTAGG GGCCTGGGG GCACCTTGGG GGTTCGACAC AGGATACTTC  
1205 E T G A H H V C F Q N S A L V I P H P R T P P G T P K L C P H K

3801 CGGTGTGGG TGAGGTAGTG AAGGGGGG TAGTGTGTG TGGAACCCAG AACGGACGC CGGTGCTTGG AGGGTTCTT AAATTATATT TAAAAAGTA  
CCACACCCC ACTCCATCAC TTTTCCCGCC ATCAACCAAC ACCTTGGCTC TTGCTGCG GCCAGAAC TCCTCCAAAG TTAATATAA ATTTTTCAT  
1238 G C G V R O O K G R O L V V E P R N G R R C L E G F L N Y I O K S N

3901 ACTTTTGTG TAAATAAAG AAATGGAC GTGTCCAGC TCCAGGGGTA AAAAAAAA AAAAAA  
TGAACACAT ATTATTTC TTTTACCTG CACAGGTG AGTCCCAT TTTTTTTT TTTTTTT  
1272 P L Y K O K K H G R V P A P G V K K K K K

35/35

FIGURE 9                      ARNILVNSNLVCKVSDFGLSRFLEDDTSDPTYTSALGGKIPMRWTAPEAIQYRKFFASAS

FIGURE 10                    NVLVKSPNHVKITDFGLARLLEGEKEYNADGGKMPIKWMALECIHYRKFTHQ

FIGURE 11                    NCMLAGDMTVCVADFGLSWKIYSGATIVRGCAKLPVKWLALGSLADNLYTVHS

FIGURE 12                    NCLVGKNYTIKIADFGMSRNLVSGDYY

FIGURE 13                    TRNILVENENRVKIGDFGLTKVLPQDKEYYKVKEPGESPIFWYAPESLTSLSFVASD

FIGURE 14                    ARNILVNSNLVCKVSDFGMSRVLEDDPEAAVYTRGGKIPIRWTAPEAIYRKFTSASD



## INTERNATIONAL SEARCH REPORT

PCT/US 93/00586

International Application No

<b>I. CLASSIFICATION OF SUBJECT MATTER</b> (If several classification symbols apply, indicate all) <sup>6</sup>		
According to International Patent Classification (IPC) or to both National Classification and IPC		
Int.Cl. 5 C12N15/12; C12N15/54; C12N9/12; //C12Q1/68, C12N15/11		
<b>II. FIELDS SEARCHED</b>		
Minimum Documentation Searched <sup>7</sup>		
Classification Systems	Classification Symbols	
Int.Cl. 5	C12N ; C12Q ; C07K	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched <sup>8</sup>		
<b>III. DOCUMENTS CONSIDERED TO BE RELEVANT<sup>9</sup></b>		
Category <sup>10</sup>	Citation of Document, <sup>11</sup> with indication, where appropriate, of the relevant passages <sup>12</sup>	Relevant to Claim No. <sup>13</sup>
A	NEURON vol. 6, no. 5, May 1991, pages 691 - 704 LAI, C. & LEMKE, G. 'An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system' see the whole document --- -/--	1-7
<p><sup>10</sup> Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"A" document member of the same patent family</p>		
<b>IV. CERTIFICATION</b>		
Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report	
07 JUNE 1993	02 -07- 1993	
International Searching Authority	Signature of Authorized Officer	
EUROPEAN PATENT OFFICE	ANDRES S.M.	

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claims No.
A	<p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA vol. 88, no. 23, 1 December 1991, WASHINGTON US pages 10411 - 10415 HOLTRICH, U. ET AL. 'Two additional protein-tyrosine kinases expressed in human lung: fourth member of the fibroblast growth factor receptor family and an intracellular protein-tyrosine kinase'</p> <p>---</p>	1-7, 13-15
A	<p>GENE vol. 110, no. 2, 15 January 1992, AMSTERDAM NL pages 205 - 211 BRÄUNINGER, A. ET AL. 'Isolation and characterization of a human gene that encodes a new subclass of protein tyrosine kinases' see the whole document</p> <p>---</p>	1-7, 13-15
A	<p>PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA vol. 87, November 1990, WASHINGTON US pages 8913 - 8917 PARTANEN, J. ET AL. 'Putative tyrosine kinases expressed in K-562 human leukemia cells' see especially clone JTK10</p> <p>---</p>	1-7
A	<p>MOLECULAR AND CELLULAR BIOLOGY vol. 9, no. 4, April 1989, WASHINGTON US pages 1587 - 1593 HAO, Q.-L. ET AL. 'Isolation and sequence analysis of a novel human tyrosine kinase gene' see the whole document</p> <p>---</p>	1-7, 16-18
P,A	<p>WO,A,9 214 748 (AMERICAN CYANAMID COMPANY) 3 September 1992 see the whole document</p> <p>-----</p>	1-12

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 93/00586

**Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:  
obscurities.  
see additional page
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/210

OBSCURITIES

Claims 2,3,5 and 7 as filed in the application are in contradiction with the description. Indeed, they are dependent on claims 1,4 or 6 which are related to DNAs or proteins of human MEGAKARYOCYTIC origin, whereas they claim some sequences which are not from such an origin, if taking reference to page 8 of the description. Therefore the search of these claims has been made independently of claims 1,4 and 6.

Claims 14 and 15 relating to the nucleotide and the protein sequences of SAL-D4 are filed as being dependent on claim 10 which relates to SAL-S1. They have been interpreted and searched as depending on claim 13.

Claims 17 and 18 relating to the nucleotide and the protein sequences of LpTK 3 are filed as being dependent on claim 14 which relates to SAL-D4. They have been interpreted and searched as depending on claim 16.

Claims 20 and 21 relating to a DNA expression vector and a cell transformed therewith, are filed as being dependent on claim 17 which relates to a protein (LpTK 3). They have been searched as depending on claim 19.

Finally, claim 2, point a) relating to SAL-S1 refers erroneously to SEQ ID 7 which describes SAL-D4. It has been interpreted as referring on SEQ ID 5 and 17).

US 9300586  
SA 69794

**07/06/93**

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9214748	03-09-92	EP-A- 0536350	14-04-93
-----			